

No gas, please, in the Middle East

Last week's conference on chemical weapons was ruffled by the flawed argument that Israel's adversaries need them to counter Israel's nuclear pretensions.

PUBLIC declarations of good intentions are cheap, decisions to put them into practice much more costly. This familiar truth seems to have been amply confirmed at last week's conference on chemical weapons (see page 199), held in Paris under the now-unfamiliar auspices of UNESCO. The 149 governments at Paris seem readily to have agreed that the best way to avoid the use of chemical weapons is to abolish them, but their joint declaration is deficient in two important ways. First, it fails to say much about the issue that has impeded the negotiations at Geneva for the past eight years, that of the verification of an agreement not to manufacture gases of the kind that can be loaded into munitions, although such a general conference cannot be blamed for that. Second, it fails to specify with clarity the dangers of the precedent set by Iraq during its recent war with Iran, when phosgene or a near relative seems to have been used against troops and civilian populations. Instead, it managed to give general currency to the notion that chemical weapons are a kind of natural counterpoise to nuclear weapons and should be legitimized as such.

That doctrine is profoundly mistaken, but seems to have been made with some force by Middle East states other than Iraq. So long as Israel has nuclear weapons (a plausible but untested assumption), is it not just that its potential adversaries should have access to chemical weapons as a kind of counterpoise? The argument is seductive, but false. The truth is that chemical weapons are not, except in the most symbolic sense, a counterpoise to nuclear weapons. Instead, they are a quite separate nuisance of a most serious kind.

Suppose that Israel is indeed a cryptic nuclear power, with a stock of more than 50 nuclear weapons — so much is the plausible inference from the harsh treatment by the Israeli courts last year of the technician from the Ramona nuclear complex who spilled the beans about plutonium extraction on the site. What use could Israel make of these weapons, which of necessity have not been tested? If Israel were in open conflict with its neighbours, and were so much in danger of defeat that the old Israeli nightmare of being driven into the sea seemed a reality, it is in principle possible that it could save itself from extinction by threatening to use nuclear weapons against the capitals of its adversaries. Earlier use of nuclear weapons, or the threat thereof, would do more harm than good to the Israeli cause — previously friendly governments would be compelled to withdraw support, and worse. But as things are, the possibility that Israel might be extinguished as an independent state would bring friendly governments rallying round long before Israel would rationally need to threaten nuclear retaliation. Thus Israel's nuclear weapons have a military function only in the extreme case that it has lost all its friends and is about to lose a terminal conflict. Militarily, Israel's nuclear weapons are a kind of national suicide pill — but their more immediate political utility is that of helping to cement deals with equally threatened governments.

Chemical weapons are a different kettle of noisome fish. Little is known of the value to Iraq of the use of phosgene in the recent war (somebody should find out), but it is likely to have been marginal in strictly military terms. Soldiers (and civilians)

were killed in dreadful ways, but the most substantial damage to the Iranian troops, ill-equipped with protection against gas, is likely to have been psychological. Nerve gases, such as Libya is said by the United States to be preparing to manufacture, would have been more damaging, especially to the civilian population. Their military utility against well-equipped Israeli troops would again be marginal, while their large-scale use against the unprotected population of some Israeli city would bring down on the head of the user an insupportable torrent of protest and constraint. But that does not mean that the prospect that nerve gases will become more widely available is of no concern. These materials would evidently be of great value in surprise attacks against small groups of people, whether troops or not. And, given the ubiquity of the trade in arms, they would undoubtedly be used in other contexts (say, in hi-jacked aircraft) well outside the Middle East.

The argument that chemical weapons are a counterpoise to Israeli nuclear weapons thus makes no sense. That does not mean that the governments in the Middle East that have raised the issue would not be within their rights in demanding that Israel should trade its putative nuclear weapons for the political and military guarantees no doubt accompanying a peace settlement in the Middle East. It may also be prudent to accept last week's demand that the scope of the Geneva negotiations should be broadened to include at least Iraq (although, with 40 participants already, the negotiations are cumbersome enough). But the argument that chemical weapons should be counted as part of a poor country's right to self-defence is paper-thin, and should be denounced as such. To make it is dangerously irresponsible for the curiosity it will excite, not only in the Middle East. □

Industrial illogic

One of Britain's largest companies may disappear because, as in the parable of the talents, it has been over-thrifty.

WHETHER or not 1992 will see the eventual emergence of the true European common market advertised in the 1950s by the Treaty of Rome, the prospect that it might has caused turmoil in what is loosely called the electronics business. Two independent sets of motives have set off a wave of industrial reorganization. European companies, recognizing that there are more of them than the European market can comfortably accommodate, are anxious to consolidate their interests, preferably across European frontiers. Second, companies from outside the European Community are seeking to establish themselves within it, not least as an insurance against the chance that common-market Europe will chauvinistically protect its industries against competitors. The latest reorganization, mostly in Britain, also points to past errors by government and business which are all too likely to be repeated in the future.

The centre of the turmoil in Britain is the General Electric Company, or GEC, which resembles in the pattern of its activity the US company called General Electric or GE, but which is only a third as big. GEC as it is owes its existence to the convic-



tion of the Labour government of 1964 that Britain's four principal manufacturers of power generation equipment then were individually too small to compete internationally and collectively too many to survive; by coaxing and coercion, they were made to merge. The result is a business of extraordinary diversity — it will make you a domestic refrigerator, a computer-aided tomography scanner or a military radar — which has been kept intact and profitable only by the talented administration of its chief executive, Lord Weinstock.

But the conglomerate has failed to make a distinctive mark in any single field; Marconi, technically GEC's pride, conspicuously failed last year to develop an early-warning airborne radar to the British defence ministry's specification (the business went to Boeing instead). Another defect is that GEC first accumulated and has since kept liquid reserves of £1,500 million. But should not companies be thrifty, salting away money for a rainy day? Not if the implication is that they are so dismayed by the difficulties of making money in their own businesses that they ask the banks to try instead. The truth is that GEC has grown too slowly, almost certainly because it has put too little of its own resources into self-financed (as distinct from military) research and development.

Now there is a shake-up on the way. Three years ago, GEC awoke from its cautious lethargy and tried to buy the British telecommunications manufacturer Plessey, but was thwarted by the defence ministry's argument (to the Monopolies and Mergers Commission) that competition among defence contractors would be thereby curtailed. Last year, GEC instead merged its own telecommunications business with Plessey's, but then, this year, cheekily combined with Siemens of West Germany in a second attempt to buy Plessey. That has brought the roof down on its head. Plessey is working hard to recruit support for an attempt to take over GEC, which has been stampeded into merging many of its business with the comparable European businesses of GE of the United States.

The company's misfortunes are partly self-inflicted, partly caused by successive British governments. GEC, like Siemens and GE, began life with the commercialization of electricity a century ago, and became a conglomerate because electricity turned out to have diverse applications. But, for twenty years, GEC has been managed as a collection of separate businesses. That should have given it an advantage over its competitors in being able easily to hive off its activities to others prepared to back their conviction that they could do better with them by paying above the odds for the opportunity. But staying big has been counted more important.

Successive governments have fortified GEC in its complacency by their shortsightedness over defence research. The late Sir Ieuan Maddock, when chief scientist at the then Ministry of Technology, was forever asking why the lion's share of British governments' research and development expenditure should go for aerospace and associated electronics, mostly for the military. The decision to oppose the Plessey merger three years ago was falsely predicated on the belief that the British services must always be equipped by British manufacturers. It would have been wiser if the government had been pressing for a European understanding that the rules of European free trade should also apply to defence procurement. That has now become an urgent need.

Meanwhile, the air is filled with talk of what is called "industrial logic", mostly from people whose skill is the management of money. The calculation in 1964 that the British market could not support four power-generation manufacturers may have been correct, but "industrial logic" would have required that the four should compete with each other for survival; the outcome might have been a manufacturer capable of winning business from elsewhere as well as Britain. In other fields within GEC's present ambit — consumer electronics, for example — there can have been no danger of over-competition a quarter of a century ago. Putting nearly all of Britain's eggs into an oversized com-

pany has allowed West German, Italian, Scandinavian and Japanese manufacturers to win handsome prizes. The whole sorry episode should be a reminder to those who talk of industrial logic that success in industry is an empirical, not an abstract, business. The winners are those who succeed in competition with each other. Selling prices matter, but so do quality and performance, all of which are best ensured by intelligent research and development. □

Research in hazard

Last week's US budget promises generosity to science, but only the Congress can decide.

LAST week's valedictory budget from President Ronald Reagan is no less what had been expected of him than the intended tear-jerker of a speech he delivered to the United States a few days afterwards. For the budget, like the speech, has an olympian quality of detachment. Defence spending would increase (but modestly) and, within that, spending on the Strategic Defense Initiative would also be increased. The National Aeronautics and Space Administration would have the funds to build its space station, the Department of Energy would be able to make a start on the Superconducting Super Collider (and the clean-up of its military reactors) and general support for basic science (through the National Science Foundation) would also be increased. So how does the federal deficit affect these plans? On paper, it seems simple. If the cost of social programmes, Medicare and the like, is reduced by \$30,000 million, the requirements of Gramm-Rudman can be satisfied for another year. The likelihood that a Congress controlled by Democrats will accept this neat solution may be small, but that will be for President George Bush to worry about when he takes over this Friday.

The danger in this process is that it puts the eminently desirable elements of the outgoing administration's programme in hazard to the uncertainties of a political process that has not yet begun. The next step is for the new president to say how he would amend the Reagan budget. Only then will the Congress begin to say how it will require the package to be changed. It would not be surprising (but par for the course) if the exact balance of the budget for the year beginning on 1 October is not known until after it has begun. In the circumstances, it is natural that the presidents of the US national academies (science, medicine and engineering) should have given the new administration their advice on how to manage science.

The Reagan administration's calculation that more spending on basic science would, in the long run, contribute to the competitiveness of US industry and the general wellbeing of society is correct, and in that spirit much has been done to increase, for example, the budget of the National Science Foundation. But some of the more expensive items in what is generally considered to be the science budget, the planned space station for example, are hardly science at all, while there is no known rational basis for telling in advance whether projects of that kind are more likely to yield economic benefits. A stronger science advisory apparatus at the White House could help to open such questions to discussion. Dr Frank Press's argument last year that the time has come when the scientific community in the United States must help to determine priorities is entirely apt.

But will it not be bad for science that one part of the community should be seen to be sceptical of another's pet projects? As in Britain a decade ago, the reality of government support for research and development is determined both by the sums of money set aside in budgets and by the value of what the money will buy when the time comes for spending it. No purpose will be served, in the United States, if promised spending is undermined by rampant inflation or by one of the other consequences of a continuing federal deficit on the scale bequeathed by Reagan to Bush. □

Scandal over Nazi victims' corpses rocks universities

■ Television programme opens old wounds

■ New mood among medical students

Munich

ISRAEL last week protested to the West German government after a television company broadcast claims that tissue samples and skeletons from the corpses of victims of Nazi-era executions are being used for teaching purposes at West German medical schools.

The Israeli Religion Minister demanded that West German Chancellor Helmut Kohl immediately deliver all body parts from Nazi victims for a proper burial, although it has not been established that Jewish victims are involved. Kohl in turn has called on the universities to stop using any such remains and added that these practices "should have ended forty years ago". The medical schools at Tübingen and Heidelberg are at the centre of the row. Both are looking into the allegations although it is not expected that the charges will be conclusively proved in more than a few cases.

The allegations provoked a storm in West Germany when broadcast on the ARD television network. Although it is well known that, during the Nazi era, corpses for medical education were procured by universities from nearby execution sites, nobody seems to have asked about the origins of the slides and skeletons now being used in anatomy classes.

The practice of using the corpses of executed criminals began before the Nazis gained power, but the number of executions rose dramatically after 1933. The anatomy institute at Tübingen alone received 1,077 bodies from the execution

site in Stuttgart between 1933 and 1945.

Tübingen University officials said they "had been lulled into a false sense of security" and believed that all remaining body parts from Nazi victims had been disposed of just after the war in a dignified manner. Both Tübingen and Heidelberg promptly began investigations but spokesmen for both universities pointed out that it will be virtually impossible to determine the origins of organs or individual bones in their medical collections.

Officials at Tübingen found four slides that had been prepared from two corpses of Nazi victims who were executed for political reasons. One was a woman of Polish extraction, the other a man presumed to be a German. In Heidelberg, three slides (out of about 1,500 in the collection) were found dating back to 1941 and 1943. The names of the people from whose corpses they were taken were not listed on the slides, but the indication 'decapitatus' indicates that the people may have been victims of Nazi persecution. A skull was also found that had been preserved by the 'preparator' of skeletons employed in the anatomy department from 1937 to 1946. All four items were removed from the department's collection and will be cremated.

That the row has erupted now may be indicative of changing attitudes. The current generation of medical students is demanding a deeper evaluation of the Nazi legacy in West German medicine. One result may be further scandals to come.

Steven Dickman

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Breaking ranks

THE pay dispute between academics and management in Britain's universities took a surprise turn last week when the head of one university made a pay and promotions offer to staff in an attempt to avert industrial action which is disrupting university examinations throughout the country.

The principal of the University of St Andrews, Professor Struther Arnott, offered all academic and related staff a pay increase of 5 per cent from January to March this year and from April to March 1990 conditional on any new national pay scales negotiated. He also offered to make 30 promotions over the next three years, at least 20 to the level of senior lecturer. The offer must be sanctioned by the university's governing body, next month.

The Committee of Vice-Chancellors and Principals (CVCP) was surprised by the move and fears that it may jeopardize the universities' negotiations with the government for extra money to finance salary increases for 1989-90. The Association of University Teachers will not accept the offer, arguing that a local solution will not solve a national problem. But it hopes the move will put pressure on the CVCP to speed up national negotiations. C.McG.

LA cools it

BIOTECHNOLOGY gadfly Jeremy Rifkin has turned his attention to global warming, with the result that before too long there should be pockets of green dotting the asphalt landscape of Los Angeles. Last week, Rifkin joined Los Angeles Mayor Tom Bradley at a press conference at which Bradley announced that the city would plant 2-5 million trees to reduce carbon dioxide emissions and break up "heat islands" that raise the local temperature.

The tree-planting exercise is an initiative of Rifkin's Global Greenhouse Network, a group made up of international public interest groups and legislators. Rifkin expects Los Angeles to be the first of a number of cities around the world to adopt the tree-planting scheme. J.P.

Czech explosives

THE Czechoslovak government, evidently embarrassed by reports that its explosive "Semtex" was responsible for the destruction of the Pan-American aircraft over Scotland last month, last week proposed an international convention on the certification and detection of explosives. The offer, made during last week's conference in Paris (see page 199), immediately preceded the visit to Britain of a Czechoslovak team to assist in the forensic investigation of the disaster. The proposed convention is intended to regulate the manufacture of explosives for legitimate purposes and would include guidelines for their detection and the prevention of their unauthorized disclosure. V.R.

Associated Press

Work continues at Dumont d'Urville in Antarctica (see last week's Nature), where France is constructing a new airstrip. Greenpeace protestors are still occupying the southern end of the airstrip.

Japan's ERATO programme found to be working well

Tokyo

JAPAN'S Exploratory Research for Advanced Technology (ERATO) programme was set up in 1981 as an experiment to try to break away from the rigid hierarchy of universities and stimulate the kind of creativity that many Japanese researchers think flourishes in the West. Last month, in its first-ever evaluation by a group of foreign researchers, many parts of ERATO pass with flying colours.

The greatest praise was reserved for the Quantum Magneto-Flux Logic Project (1986-91) led by Eiichi Goto of Tokyo University, which is exploring the possibility of making ultra-fast computers using magnetic quanta. John Rowell of Bell Communications Research, who assessed the project, said he was amazed that such a "sense of rebellion" could be fostered within the confines of Japan's existing industrial and university structure.

He believes the project has achieved success, as exemplified by the demonstration of a variety of logic functions using single flux elements composed of Josephson junctions. Such progress could not have been made in the United States, he says, where expertise in Josephson junction technology has been allowed to "evaporate".

ERATO, which is run by the Science and Technology Agency's Research and Development Corporation, has financed 18 projects, most of them high-risk research with some potential for application. Two or three new projects are begun each year and funded for five years at \$2-3 million per year. Except for the project head, no member of the 15-20 person project teams may be over 35 years old. Researchers are drawn from industry and universities and include several foreigners.

The new evaluation comes from the Japan Technology Evaluation Program, which is supported largely by the US National Science Foundation. The Perfect Crystal Project (1981-86) headed by Junichi Nishizawa, Tohoku University, which developed improved methods for growing perfect crystals of silicon and gallium arsenide, was another project judged "very successful". Edward Wolf, director of the National Nanofabrication Facility at Cornell University, said he was particularly impressed with the effective coupling of the project with industry.

Among the biological projects, the Bioinformation Transfer Project (1983-88) headed by Osamu Hayaishi, Osaka Bioscience Institute, which examined prostaglandins in the central nervous system and their role in biological information transfer, is given high marks by Dale Oxender, director for molecular genetics at the

University of Michigan. Seven patents have been filed and the results may have important applications in the pharmaceutical industry, Oxender says.

Seen as successful, but less imaginative, was the Superbugs Project led by Koki Horikoshi of RIKEN, which aims at finding commercial applications for microorganisms that can thrive in extremes of pH, temperature or salinity. The Bioholonics Project (headed by Denichi Mizuno, Teikyo University) and the Molecular Dynamic Assembly Project (H. Hotani,

Kyoto University) were seen as "comparable to similar efforts in the United States".

Nearly all evaluators noted that ERATO gives project leaders considerable freedom to spend money as they wish, in contrast to the bureaucratic regulations surrounding financial support from the Ministry of Education, Science and Culture, But university researchers who accept ERATO funds are almost invariably "cutoff" from Ministry of Education funds unless they are "politically powerful". And Rita Colwell, director of Maryland Biotechnology Institute, drew attention to the total lack of female scientists in ERATO and the lack of leadership opportunities for female scientists in Japan.

David Swinbanks

AAU issues guidelines on dealing with scientific fraud

Washington

NEW guidelines intended to help universities to find ways to "distinguish fraud from the honest error and the ambiguities of interpretation that are inherent in the scientific process" were issued last week by the Association of American Universities (AAU), a body that represents most big US research universities.

The guidelines come as public concern over scientific misconduct is increasing, carrying with it the possibility that the US Congress will introduce strict new legislation to control research. A key part of the thinking behind the guidelines is that "institutions ought to be held responsible

been reached, the institution should discharge its responsibilities to "the public, the sponsors of research, the scientific literature, and the scientific community".

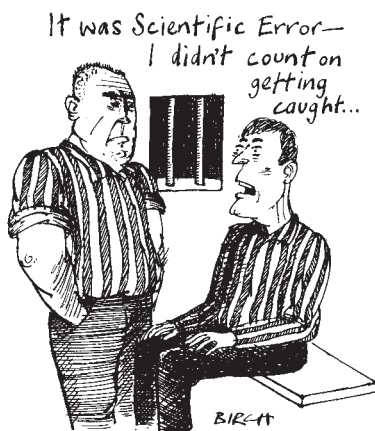
The guidelines take up two of the commonest criticisms of misconduct investigations: that they take too long and that they can cause greater harm to the complainant than to the guilty.

A preliminary inquiry should be completed within 30 days and a full investigation, if warranted, within 120 days, although an institution may "acknowledge formally" that the nature of the case makes it impossible to meet the deadline. The guidelines acknowledge that "younger, less senior people are particularly vulnerable" when making accusations of misconduct, and urges that complainants be protected.

Little is made of another common criticism, that internal investigatory committees are inevitably under pressure to save face for the institution involved.

The new guidelines have been sent to all of AAU's 54 member institutions who will decide for themselves if they wish to use them as the basis for dealing with research fraud. AAU member institutions between them receive 60 per cent of all research grants given out by the National Institutes of Health (NIH), but the PHS, which is responsible for NIH, has yet to complete its own guidelines. An official notice of "proposed rule making" was published four months ago, more than two years after initial guidelines were produced. The final regulations are unlikely to be ready before May.

Alun Anderson



for the conduct of their faculty", according to Robert Rosenzweig, AAU president.

The new guidelines are simpler and stricter than draft guidelines issued by the Public Health Service (PHS) in 1986 and by the National Science Foundation in 1987. The guidelines do not accept that the resignation of a faculty member can close the books on a misconduct investigation, even though that action may provide an institution's simplest way to avoid public scandal. And once a conclusion has

Erratum

Contrary to a News item in *Nature* 337, 7; 1988, it is Johann Deisenhofer who is now in Texas (at the Howard Hughes Institute, University of Texas Southwestern Medical Center, Dallas) and Hartmut Michel who has remained at a Max Planck Institute (in Frankfurt). □

OECD report itemizes risks of renewable energy production

Paris

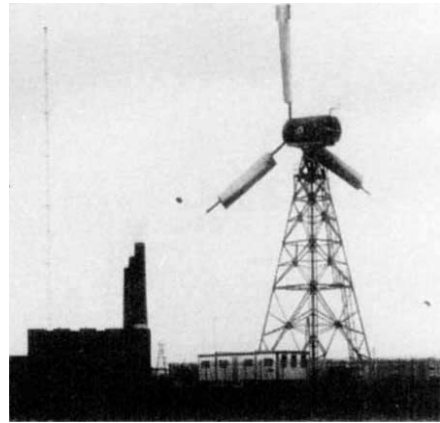
ENVIRONMENTALISTS have been campaigning for years to alert governments to the long-term risks of burning fossil fuels for energy production, but the so-called 'clean' alternatives of wind, wave, solar power and geothermal power are not without serious hazards, according to a report from the Organisation for Economic Cooperation and Development (OECD).

The report takes the "important positive aspects" of renewable energy use as given. But while the negative impacts of fossil fuels or nuclear energy may be long-term and planet-wide, those of renewables tend to be more localized and shorter-term. In most cases, they arise because renewable sources convert low energy densities and thus require huge collector areas. Hydroelectric power, which accounts for 21 per cent of energy production in OECD countries (1978 figures), like wind power, is not directly polluting, but may render large areas unusable for agriculture or recreation and can destroy local ecosystems.

Biomass, especially firewood, which accounts for about 20 per cent of energy produced in developing countries can, however, have serious environmental impacts. Apart from deforestation, with its risks of soil erosion, burning firewood can produce airborne, inhalable particles that may be carcinogenic, as well as greenhouse gases. Large-scale waste incineration, on the other hand, is only as clean as the material burned. If the waste stream contains plastics or chlorinated products, highly toxic and often halogen-

ated compounds may be produced, including dioxin and hydrochloric acid.

If passive solar energy presents few risks, apart from the build-up of radon due to improved insulation, active solar energy panels contain potentially dangerous chemicals, such as biocides, rust-inhibitors and antifreeze that could pose a



Good news but not if you are too close. . . This was the first wind-turbine from Britain's CEBG.

problem for disposal on a large scale, says the report.

Geothermal energy also poses risks. Hydrogen sulphide and benzene are almost always associated with high-temperature geothermal fluids in large volumes and there is also an unknown risk that reinjected geothermal fluids could induce seismic activity — and most geothermal sites are in seismically active regions.

Peter Coles

Environmental Impacts of Renewable Energy. The OECD Compass Project. OECD 1988.

Indo-French agreement stresses vaccines

New Delhi

A MAJOR Indo-French agreement on medical research is due to be signed in February on the eve of the visit of French President François Mitterrand to New Delhi to open the year-long festival of France in India. The agreement will be between the Indian Council of Medical Research and the French Health and Medical Research Institute (INSERM).

Seven research areas will be covered: fertility and human reproduction, ophthalmology, nutrition, public health, epidemiology, virology, immunology and bio-engineering. Collaboration will be on equal terms and will include exchange of scientists, sharing of information and the initiation of research projects. The agreement will continue the trend towards increased collaboration between India and France.

An Indo-French Centre for Advanced Research set up in New Delhi in 1987 has

already given shape to 14 joint projects. It will soon have a building of its own, constructed by France and given to India as part of the festival.

Another collaborative venture for the manufacture of vaccines is taking shape. It involves the construction of a plant to mass produce vaccines against measles, polio and rabies, and a multiple vaccine, in Gurgaon, on the outskirts of New Delhi, by the Institut Merieux of Lyons. The plant will at first meet the demands in India and later be expanded to cater to neighbouring countries.

Collaboration on vaccine production is important in the context of India's mission to immunize all children against six diseases by 1992. Another plant for producing 100 million doses of oral polio vaccine is being set up at Bulandshar, 140 kilometres from Delhi, in collaboration with the Soviet Union.

K.S. Jayaraman

Reactor restarts sans repair

Paris

THE French prototype fast-breeder nuclear reactor, Superphénix, is to be restarted after a 20-month shut-down, even though its leaking fuel-rod transit chamber has not yet been modified (see *Nature* 332, 384; 1988). The news, published in the government archive *Journal Officiel* last week, follows an announcement by the Industry Minister, Roger Fauroux.

The decision was greeted with applause by the 200 employees at the reactor's Creys-Malville site, near Lyons. But ecology action groups in France and neighbouring Switzerland are outraged and will now seek legal help to try to shut down the reactor permanently.

Superphénix was closed in March 1987 when liquid sodium coolant was found to be leaking from a storage and transit chamber. Engineers from Electricité de France (EDF), the nationalized utilities company that runs Superphénix, found fissures in the inner lining of the chamber. They proposed replacing the sodium coolant with inert argon gas and then running the reactor without the transit chamber inner lining in place. As the core is already charged with fuel, says EDF, the reactor can be run safely for about a year without needing to use the transit chamber. In the meantime, the chamber can be modified.

Fauroux was planning to restart the reactor on 14 January. If there are no incidents, electricity generation could begin again in March.

Critics feel that it is irresponsible to restart the reactor with one of its original safety devices out of action. Should an accident occur within the core, the transit chamber is the only place where fuel rods can safely be stored. They argue that Superphénix is not actually needed for electricity generation and is essentially a full-scale simulator for engineers.

Environmentalists say the government's action is undemocratic. "The decision was designed to have a psychological impact, to demoralize the opposition", says René Longet, director of the Swiss Association for the Protection of the Environment. "Why cannot the French allow open debate on nuclear energy, like other European countries?", he asks. France appears increasingly isolated now that both Britain and West Germany have effectively abandoned their fast-breeder programmes. But France, too, is under internal pressure, mostly economic, and has shelved plans to build a sister to Superphénix. The Centre à l'Energie Atomique, however, remains adamant that it is essential to gain technical expertise at Creys-Malville.

Peter Coles

Fuzzy computing felt to be the next step in Tokyo

Tokyo

JAPAN's Ministry of International Trade and Industry (MITI) has persuaded a vast consortium of Japanese companies to invest millions of dollars into establishing a new centre for research and development on fuzzy computing.

The decision to set up the facility, called the Laboratory for International Fuzzy Engineering Research (LIFE), was made at a meeting of MITI officials and industry representatives on 11 January. According to a report in the *Nihon Keizai Shimbun* (an economic newspaper), 42 companies — including Toshiba, Nippon Steel, Hitachi, and Toyota — will each contribute ¥100 million (\$800,000) to LIFE. The project, not run directly by MITI, will be chaired in alternating two-year shifts, beginning with Hitachi. The centre, scheduled to open in April in Kannai near Yokohama and expected to operate for six years, will be staffed by some 30 researchers drawn from industry and will conduct joint research with government national laboratories (for example, Tsukuba's Electrotechnical Laboratory) and universities (specifically, Tokyo Institute of Technology and Kyushu Institute of Technology). MITI also plans to collaborate with foreign universities and research associations in countries such as France, the United States and West Germany. Patent rights will be held by LIFE and licensed to non-members, but fuzzy algorithms will be made freely available.

Fuzzy computing is based on fuzzy set theory, first presented in 1965 by Professor L.A. Zadeh, a US mathematician at the University of California at Berkeley who is now chairman of the International Fuzzy Systems Association there. Whereas conventional digital computers operate on the binary system, fuzzy computers can form conclusions from imprecise data (such as 'faster' or 'not so hot'), treating such input as having a value somewhere in the continuum between zero and one. Proponents of fuzzy theory hold that these characteristics will allow fuzzy computers to process information in a manner analogous to the brain.

Considerable research on fuzzy computing has already been carried out in Japan, and fuzzy programming on conventional computers is used to control train speed in a subway in Sendai, a city north of Tokyo. At Kumamoto University, Associate Professor Takeshi Yamakawa has developed a fuzzy micro-computer chip set in collaboration with Omron Tateisi Electronics Company. Of the two chips, the 'rule chip' draws inferences from indistinct data; the 'defuzzifier

chip' converts inferred conclusions into analog numerical values. In September 1987, the US National Aeronautics and Space Administration (NASA) asked Yamakawa if it could use his technology in controlling attitude and docking in the space shuttle. Yamakawa is said to have agreed on condition that the chips be used solely in civilian applications. NASA is said to have complied. Sometime this spring — after NASA completes its evaluation — Yamakawa intends to commercialize the device.

"What we are interested in falls into three categories", says MITI spokesman Taizo Nishikawa, deputy director of the Electronics Policy Division at the Machinery and Information Industries Bureau, "fuzzy control, such as the Sendai system; fuzzy information processing; and fabrica-

tion of fuzzy hardware — for example, fuzzy chips — that will be used to evaluate our basic research in the first two categories."

"There is a very broad range of applications for fuzzy computing", continues Nishikawa. "Industrial applications include image processing, plant management, and robotic control. Social applications include medical diagnosis, weather forecasting and aeronautics control."

The scope of the industries from which member companies come is quite broad, as are their specific interests in fuzzy applications. In addition to the steel, electronics and automotive industries represented by the companies listed earlier, members include computer software houses, chemical companies, electrical appliance and consumer electronics manufacturers, transportation companies, an electric power utility, and — very unusual for a MITI technology project — a Tokyo securities firm.

Stuart M. Dambrot & David Swinbanks

This year's prize season gets under way

London

ROBERTO Poljak of the Pasteur Institute in Paris, Walter Schaffner of the University of Zurich and Gregory Winter of the Medical Research Council's Laboratory of Molecular Biology in Cambridge are to share the 1989 Louis Jeantet Prize in Medicine, awarded for outstanding biomedical research in Western Europe. All three are selected for their contributions to immunology, and each will receive about one-third of the £0.75 million prize, plus £30,000 pocket money.

Poljak is cited for his contributions to solving the three-dimensional structure of antibodies, and will use his prize money to improve the computing and graphics power available to his group, and for new area detectors for his X-ray diffraction equipment. He hopes to solve the structure of a lysozyme idiotope-anti-idiotope complex. Schaffner will invest in an automated cell-microinjection apparatus and a fluorescence-activated cell sorter, both to enable his group to study the relevance of DNA methylation to the expression of genes, particularly the antibody genes for

which he is cited.

Winter is rewarded for work on engineering antibody molecules, and hopes to finance a project he has been hoping to undertake for some years — to design an antibody molecule to bind specifically to an antigen of known three-dimensional structure. Winter has also won the 1990 Emil von Behring prize given by the Philipps University of Marburg. He wins the £6,000 prize for his contributions to the molecular biology of the development of human monoclonal antibodies.

The 1989 King Faisal International Prizes, each worth £0.5 million, have also just been announced. The medicine prize is shared by Robert Edwards of the University of Cambridge and Luigi Mastroianni of the University of Pennsylvania School of Medicine for work that has made human *in vitro* fertilization possible. Theodor Hänsch, who has recently moved from Stanford University to Munich University, and Ahmed Zewail of California Institute of Technology, share the science prize (this year physics) for their laser studies.

Peter Newmark

Poljak (left), Schaffner and Winter: the immune system pays handsome dividends.



Chemical arms exports embarrass Bonn

Bonn

ON 11 January, as West Germany and 148 other nations resolved at a Paris conference to eliminate chemical arms worldwide, the first hard evidence began to appear linking West German companies to the alleged Libyan chemical weapons factory at Rabta. By now, as many as 30 firms have been implicated, although the West German government still insists that it has not seen evidence "admissible in a court of law." One day earlier, West Germany had announced new measures to restrict exports of "sensitive" goods of a military, nuclear or chemical nature. The restrictions came in response to a December scandal in which at least two West German firms are suspected of having delivered nuclear technology to Pakistan, India and South Africa in possible violation of the nuclear non-proliferation treaty.

The revelations about West German participation in the Rabta plant embarrassed the Bonn government. West Germany had been under pressure from the United States to investigate exports to Libya by Imhausen-Chemie GmbH, a company based in the Black Forest. A first investigation by tax authorities in Freiburg in early January cleared the firm.

Chancellor Helmut Kohl refuted the US charge, saying on 10 January that the United States had "put the Germans in the dock without giving us a chance to see the evidence." The next day, the news magazine *Stern* broke the news that the West German customs service had evidence against Imhausen and other firms. At the end of the week, the government revealed that the Federal Intelligence Service (*Bundesnachrichtendienst*) had informed Kohl in October that Imhausen might have violated export laws. The public prosecutor's office in Offenburg has begun a new investigation.

Regulating exports has always been a sensitive matter for West Germany, whose annual exports are worth more than DM 500,000 million. The new regulations put tighter controls on the export of technology and know-how in the nuclear, chemical and munitions areas. The government also recommends increasing the penalties for breaking export laws. The maximum prison term will be raised from three years to five years and the maximum fine from DM 500,000 to DM 1 million.

The government hopes to restrict future arrangements like the alleged Imhausen-Libya deal by lengthening the list of "sensitive areas," which previously included only the Soviet bloc and countries at war. Furthermore, the government recommends prosecuting Germans overseas found to have a hand in the production of chemical or biological weapons.

The opposition criticized the government measures while calling for a parliamentary investigation of the "poison gas affair" as it is coming to be called. Bundestag (parliament) member Norbert Wiczorek (Social Democrat) said it was hypocritical of the government to oppose exports which could lead to chemical arms production only when they go outside the European Community (EC). Wiczorek said that "there are not enough teeth" in the new regulations. He said that the fines discussed were "ridiculous sums" which would be considered by violators as just another operating cost.

Wiczorek admits that tougher controls do not guarantee that undesirable exports will stop. The volume of exports is so great, and the customs and regulatory authorities so understaffed, that many products will continue to pass through undetected. Only by increasing the penalties and the number of spotchecks will

would-be violators be discouraged.

Perhaps the thorniest part of the problem comes in the area of "dual-use" technologies, such as the pesticide factory built at Samarra, Iraq, with the help of West German companies. It has now been established that the factories produced poison gas, which Iraq apparently used during its 8-year war with Iran and on its own Kurdish population. But the guilt of the West German firms remains in doubt because it is unclear if they knew how their products would be used. An investigation into the matter is in its fifth year.

Another firm, Karl Kolb AG of Dreieich, has avoided prosecution under the current export laws by challenging the laws in court. Kolb was one of the companies charged with delivering equipment for the pesticide factory to Iraq. Kolb won the first court case; the government is now appealing.

Steven Dickman

Some good news some bad from Paris

Paris

THE five-day Paris conference on chemical weapons closed on 11 January with all 149 participating nations declaring their readiness to "prevent all recourse to chemical arms, by eliminating them completely". But the declaration has already been criticized as not going far enough.

Britain, for example, would have liked concrete measures to prevent the proliferation of chemical weapons and an international commitment to punish governments that used them.

The declaration should, however, speed progress at the Geneva disarmament conference, whose 40-nation membership is trying to formulate a legally binding decree to ban the production and stockpiling of chemical weapons. The declaration asks the Geneva conference to redouble its efforts "urgently to resolve remaining problems" and to draw up a convention.

The plenary committee set up to draft the declaration had to compromise after heated exchanges among the participants. Iran, in particular, wanted Iraq to be pilloried. Syria defended possession of chemical arms so long as Israel has the nuclear bomb. Initial insistence by some other developing nations that there could be no chemical disarmament without nuclear disarmament was eventually dropped. In the final document, individual nations were not named, although the use of chemical weapons in recent wars is condemned.

There was further trouble when the South African foreign minister, Pik Botha, took the platform on 9 January. Many African delegates walked out, as well as one French delegate. But the biggest

news from the conference remains the announcement that the Soviet Union will halt production of chemical weapons immediately and will start the ten-year process of destroying its stocks. The Soviet move includes an acceptance that international inspection will be essential to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Lone voice: Pik Botha spoke but few were there to listen.

make compliance believable. But the view of the outgoing US administration is that the problems surrounding inspection are "insurmountable", given the ease with which conventional factories can be modified to produce chemical weapons.

Despite the obstacles to a complete treaty, the Paris declaration holds promise as a deterrent to further proliferation. The high profile surrounding the Paris meeting has already helped put pressure on nations such as West Germany, accused of supplying technology to Libya for chemical arms manufacture, to re-examine export controls (see above).

Peter Coles

More money for science in Reagan's last budget

Washington

PRESIDENT Ronald Reagan delivered his final budget to Congress last week, a spending plan that seeks increased support for many areas of basic scientific research. Although it is tempting to view the lame-duck budget as 'dead-on-arrival' — because Congress and the Bush administration are certain to make changes — a complete overhaul of the \$1,150,000 million budget is virtually unthinkable, and many of the proposals contained in it are certain to be approved once the new fiscal year begins on 1 October.

The Reagan years have seen substantial increases in spending for research and development. In a speech last week introducing the proposed 1990 budget, White House Science Advisor William Graham pointed out with some pride that between 1982 and 1990 "federal support for basic research will have increased by 56 per cent in real terms", reaching \$11,200 million in 1990.

Graham used his budget remarks to criticize groups such as the National Academy of Sciences which have called for a stronger science adviser in the White House. He argued that "creating a science and technology 'tsardom' to direct what is to be done from the top down would be a disaster". Graham believes the best guidance for science comes from individuals doing research, and that a government's role is to support the best research.

The new budget calls for substantially greater spending on several big science projects; the National Aeronautics and Space Administration (NASA) budget requests \$2,100 million for the Space Station, the Department of Energy (DoE) will request \$250 million for continued development work on the Superconducting Super Collider (SSC), and the National Institutes of Health will be seeking \$100 million for the human genome project.

Although the space agency's budget request is considerably more than that for the previous year, NASA administrator James Fletcher says the budget is a lean one, with no room for alteration. He does not hide the fact that the space agency had sought considerably more money, but had its request scaled down by the White House Office of Management and Budget (OMB). Although OMB approved a request for new funds for CRAF, the Comet Rendezvous/Asteroid Fly-by mission and its sister spacecraft Cassini, a planetary mission to Saturn, many prized projects have fallen by the wayside or been cut severely. There is no money for development of shuttle C, a cargo-carrying version of the space shuttle, and the Pathfinder advanced technology pro-

gramme has been cut by more than half. In addition, OMB has decreed that \$208 million must be raised from the private sector to cover expected costs, an ambitious target with which the agency is not happy.

The requested rise of \$927 million for defence activities is an indication of the Energy Department's current preoccupation with getting the nuclear weapons programme back on track and bringing its ageing facilities into compliance with environmental regulations. Environmental safety and health programmes will also request significantly more money — \$124.6 million in 1990 — but that is only a tiny fraction of the total cleanup cost at weapons facilities, estimated to be at least \$50,000 million.

In addition to requesting support for the SSC, DoE's budget also seeks construction funds for a 6–7 GeV Synchrotron Light Source at the Argonne National Laboratory in Illinois and the Compact Ignition Tokamak at Princeton, New Jersey. There is also a request for \$27.6 million for work on the human genome project, a \$10 million increase over that for 1989.

The National Science Foundation budget seeks \$2,150 million in 1990, a 14 per cent increase. Science education continues to receive emphasis, with a total of \$190 million for precollege, undergraduate and graduate science education.

The National Institutes of Health (NIH) budget is scheduled for an additional 6.6 per cent for basic research. The total budget for NIH is set at \$6,777 million, a modest 3.6 per cent rise over the current figure. This does not include a large increase in spending on research on AIDS, now listed as a separate item in the budget. The total request for federal AIDS activities within the Public Health Service is \$1,600 million.

Defence research, still accounting for nearly two-thirds of all federal spending on research, rises to \$41,024 million for the Department of Defense. The budget for the National Aerospace Plane is nearly \$300 million. The Strategic Defense Initiative request is for \$5,800 million, a \$1,800 million increase over 1989. The Defense Advanced Research Projects Agency budget request is cut by nearly \$166 million to \$1,121 million.

Congressional budget committees will begin work in earnest on the budget once they have a clearer idea from President-elect Bush of what changes they can expect. But Albert Teich, one of the authors of a recent report for the American Association for the Advancement of Science on federal spending on research in the 1980s, says there are not likely to be

big changes judging by past experience. Teich points out that in 1981, Democrat Jimmy Carter's final budget was filled with signals that the transition to a new administration would bring significant changes when Republican Reagan took over. Those signals are not present today, suggesting that changes to the Reagan budget proposals will be at the margins rather than the core of the spending plans.

Joseph Palca

UK meat research under threat

London

BRITAIN'S meat research could be severely curtailed next year by the threatened closure of the Bristol branch of the Agriculture and Food Research Council's Institute of Food Research, according to the Institution of Professional Civil Servants (IPCS), the union that represents more than half of the Bristol laboratory's 200 staff. The laboratory would be just the latest casualty of the government's policy to withdraw support from near-market research, for which it says industry should pay. The institute's director, Douglas Georgala, confirms that one of the three branches will close, but says no decision has yet been made as to which.

Many AFRC laboratories are suffering as a result of the reduction in contracts from the Ministry of Agriculture, Fisheries and Food, but the meat research laboratory has been hardest hit: its budget of £3.2 million is being halved over three years. The IPCS says efforts by the staff to become independent from government financing and secure contracts with industry are now being blocked: contracts between researchers and the Meat and Livestock Commission have been reduced from three years to two; contracts with industry and government through the LINK food-processing sciences programme will be restricted; and researchers are being prevented from submitting projects to FLAIR, the European food research programme. All these are signs, members of the Bristol laboratory believe, that their institute has been earmarked for closure.

Georgala says every effort is being made to help researchers secure contracts with outside bodies, but a temporary hold on international activities is necessary because the institute does not want to make commitments it may not be able to fulfill. He says the institute's activities will have to be cut back and is preparing a plan to implement cuts which will be presented to the AFRC in the next few months.

Closure of the Bristol laboratory will signify the end of meat research in Britain, says the IPCS, and the effect on the meat products industry, worth £8,000 million per annum, will be disastrous.

Christine McGourty

Hubble telescope delays prove a boon for some

Berkeley

ALTHOUGH frustrating, the three-year delay in launching the Hubble Space Telescope has proved a boon for at least one of the mission's scientific teams. During ground testing of the faint object spectrograph (FOS), project manager Larry Randall says "fundamental problems" were found that were causing 10–30 per cent reductions in the light-gathering capability of the blue-side and red-side spectrographs. Tests of the instrument's optical alignment revealed a 'cat's eye' where there should have been an unbroken disk of light.

Randall received permission to open the instrument in late 1986. Inside, he discovered insulation and baffle material hanging in the light beam. And pieces of black tape and a rubber glove were loose within the instrument where they could stray into the light path in a microgravity environment. Randall also found that the instrument did not have the proper optical alignment, a problem he thinks may have been caused by the unfamiliarity of the instrument's contractor — Martin Marietta — with proper alignment techniques. Lou Ripp of Martin Marietta who was involved with the FOS project, said through a spokesman that the company had made a mistake in not involving scientists more actively in the design from the start.

In the overhaul, the FOS light path was restored and the opportunity taken to fit a new detector for the red-side spectrograph. The detector uses recently declassified technology developed by a project member, Edward Beaver of the University of California, San Diego, that has been used in military night vision devices.

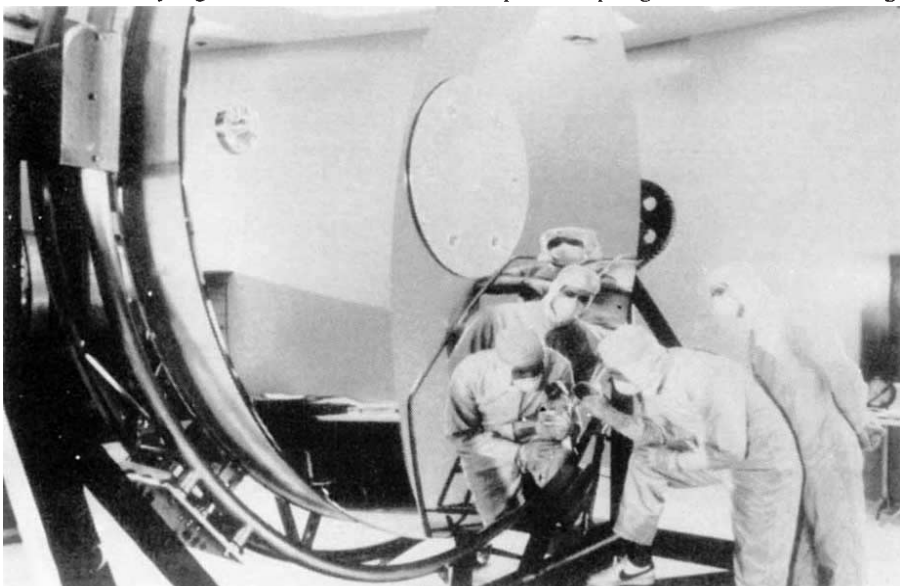
The Hubble Space Telescope is now in Sunnyvale, California, where it is in the final stages of ground testing for its December launch. Apart from the changes made to the FOS, the alterations of the telescope's other four instruments were limited to refurbishing of parts to lengthen the lifetime of instruments, and the replacement of a faulty power supply on the high speed photometer, according to Fred Wojtalik, project deputy manager.

The telescope will soon receive its new improved solar wings from the European Space Agency, modified not only to increase power levels by 15 per cent, but also to resist degradation by atomic oxygen encountered in low Earth orbit. The newly developed nickel-hydrogen batteries (*Nature* 326, 630; 1987) are still in final testing, but if they do not prove suitable Wojtalik says the craft can operate with the batteries from the original design.

The last remaining major decision is whether the telescope will travel to Florida by air or sea. NASA (National Aeronautics and Space Administration) originally planned to ship the craft through the Panama Canal on an ice-breaker. But, because the ship's busy schedule would allow only a narrow time window for the trip, mission planners proposed to fly the telescope to Florida on a C-5A transport, in an Air Force container.

The telescope will fly to its Florida launch pad in August unless next month's flight tests show that the Air Force container, which was built for other purposes, is unable adequately to support and protect the telescope, in which case it will go by ship in late spring.

Marcia Barinaga



The mirror from the optical telescope aboard Hubble, during manufacture at Perkin-Elmer. It is coated with a reflection layer of aluminium $2\frac{1}{2}$ millionths of an inch thick, protected by a layer of magnesium fluoride 1 millionth of an inch thick.

UK plans for astronomy and planetary science

London

A PLAN for Britain's research in ground-based astronomy and planetary science over the next 20 years is published this week by the Science and Engineering Council. It focuses on four projects. The main one is an extension of the MERLIN radio-interferometer, based at the Nuffield Radio Astronomy Laboratories at Jodrell Bank, by the construction of a new 32-metre radio-telescope at Cambridge. This will also enhance British participation in the European Very-Long-Baseline Interferometry Network; construction of the telescope, to begin this year, will cost £6 million.

The council also recommends that Britain seeks a 25 per cent share, at a cost of £2 million, in a new incoherent scatter radar at Spitzbergen in the Arctic and at least a 30 per cent share in a new 1-km gravitational radiation observatory. The council is seeking extra money from the Advisory Board for the Research Councils so that Britain can afford a greater share in the observatory and thus secure its siting in Scotland.

The fourth project in the plan is the construction of an 8-metre telescope optimized for both optical and infrared work. At a cost of £15 million over the next ten years, Britain would hope to have a 50 per cent share in the telescope.

The plan also identifies 17 smaller projects which may be selected according to available resources. £28 million is available for the programme until the year 2000, but the board says that development of the new facilities will require savings in existing ones; these will be made through creating more international partnerships, reducing operating costs and closing some facilities. It does not specify where the cuts might be made.

Christine McGourty

Israeli satellite stars

Jerusalem

OFEQ 1, Israel's first satellite, burnt up after re-entering the Earth's atmosphere (over the Pacific Ocean) on 14 January, 118 days after launch. The 156-kg experimental communications satellite was expected to remain in orbit for only six weeks. Professor Yuval Ne'eman, head of the Israeli space agency, attributed the extended life span to the fact that Ofeq 1 (Horizon 1) reached "unexpected" altitudes, thereby encountering fewer particles in space. He said the satellite continued to transmit data on the Earth's environment and magnetic field throughout its orbit.

Ofeq II, a more sophisticated satellite due to carry a scientific payload, is being prepared for launch within two years. With Ofeq 1, Israel became the ninth country to launch a satellite.

Lisa Perlman

Anonymous peer refereeing

SIR—Recent complaints about the flaws of anonymous peer refereeing (APR) have overlooked a crucial question: how has this system arisen? Why do authors and referees regard this practice as inherently fair, so that they never really question its ethical validity? The explanation may be an historically interesting example of the unnoticed substitution of values.

The roots of APR are likely to be found in a legitimate democratic perception of anonymous voting. Anonymous voting by, say, secret ballot, is valid when issues must be determined by anonymous majority vote, say the election of governments or departmental chairmen. Scientists as a group have a deep respect for democratic traditions. But judgement on a particular research paper is not within the domain of statistical counting and is not (or at least should not be) subjected to an anonymous peer vote. Is it possible that traditional sympathy for democratic traditions has led the scientific community spectacularly to overlook the inappropriateness of APR?

ALEXANDER A. BEREZIN

Department of Engineering Physics,
McMaster University,
Hamilton, Ontario,
Canada, L8S 4M1

SIR—I should like to add my recent experience to the letters regarding anonymous referees from A. Thyagaraja (*Nature* **335**, 391; 1988) and J. B. Wright (*Nature* **336**, 10; 1988). During the past year, I have submitted three reports of theoretical calculations in an active field of research to journals in three different countries. The results of the calculations differed in one significant respect from those generally accepted and publication was refused in all cases. I have no quarrel with this but, as none of the referees was able to produce experimental evidence for the rejection of the results, it was frustrating that all the reports were anonymous.

In one case, I attempted to correspond with the referee through the editor, who refused to pass on my letter as this was against the policy of the journal. Again, I appreciate that this policy is designed to give referees more freedom to speak their mind, but I would argue that, if referees are to remain anonymous, it should be possible for an author to communicate once through the editor, after which it is up to the referee to consign the communications to the wastepaper basket if he likes.

G.C. FLETCHER

Department of Physics,
Monash University,
Clayton, Victoria 3168,
Australia

SIR—J.B. Wright (*Nature* **336**, 10; 1988) does not believe there is a single good reason for anonymous refereeing, and asks for opinions from other readers. For about twenty years, I have signed all my reviews. My strong feeling is that non-anonymity helps to keep reviews brief, helpful and relevant. Anonymous reviews tend towards rambling essays on epistemology or, worse, a reviewer's opinion, based on unpublished observations, that a particular detail is inappropriate.

Few editors even read unsigned letters. None publishes them. It is therefore inexplicable to me why scientists accept anonymous reviews of their work. Some more specialized journals already publish critiques if they make interesting points not covered in the article itself. *Nature* might try this with some major articles, at least*. The manuscripts would be much improved, and the general reader is likely to enjoy seeing other, informed points of view. Along with Wright, I would have no confidence in a referee who balks at being identified.

There is a further benefit that even the most insecure, but overworked, scientist might appreciate. Judging from a small database, it seems that insisting on having your name on all reviews substantially decreases the number of manuscripts you are asked to review.

STEPHEN ROTH

Leidy Laboratory of Biology,
University of Pennsylvania,
Philadelphia,
Pennsylvania 19104-6018, USA

* For an example, see *Nature* **223**, 161 (1969). Editor, *Nature*.

NERC rundown

SIR—The November 1988 issue of *The Biologist* made public what has been known to the Nature Conservancy Council and its advisers for some time, that the UK Natural Environment Research Council is proposing to run down the Biological Records Centre at Monkswood, a data centre of world renown. The marine science division of NERC has already closed down the long-term biological data series for the English Channel off Plymouth — called the 'Russell Cycle' — and is planning to stop the operations of the Continuous Plankton Recorder survey of the North Sea and eastern North Atlantic early next year. Both of these data series are of world-wide scientific interest, and, like the data held by the Biological Records Centre, provide important information on the effects, deleterious or otherwise, of man's influence on the biosphere.

An inescapable conclusion is that

NERC is preparing to close down most of the existing biological recording and collection of time-series data. Such continuing studies have been criticized by research council staff and civil servants on the grounds that they are "open-ended" and prevent the more "flexible" approach required by today's government. In this context, 'flexibility' means the ability of NERC administrators to direct groups of researchers towards special topics currently deemed of popular or governmental importance. Such pop topics change every few years and research would thus proceed in a series of well-publicized short-term *ad hoc* projects. It seems that NERC officials have persuaded their council members to agree that biological recording is a low priority area unworthy of NERC support and that there is a plan in the United Kingdom to run down all biological work of environmental importance, especially in the marine field. It is time biologists started protesting loudly in public and began lobbying fellow biologists who have places on the research councils.

A. J. SOUTHWARD

Marine Biological Association,
Plymouth PL1 2PB, UK

Folk deceptions

SIR—In his review of Paul Barber's book *Vampires, Burial, and Death: Folklore and Reality*¹, Roy Porter remarks: "... this Count Dracula figure, so familiar from the movies, bears scant relation to the traditional revenant who is so powerful a presence in Central and Eastern European folk memory. Examination of recorded sightings shows the vampire was quite different ...".

As a biologist who is also interested in traditional folk culture, I have spent many years doing field work in the areas that make up Transylvania, yet I have never heard any mention of Dracula, vampires or even "revenants". And in his encyclopaedic survey of ethnohistorical concepts², T.A. Szabó Sr includes only a single reference to *drakuly* (meaning 'devil') and that in 1680.

I am inclined to think that the whole Dracula story is a projection of Western European, chiefly Anglo-Saxon, ghost tradition to a distant and mysterious land. That projection has been exploited in print and in movies, best — but not first — by Béla Lugosi, whose picture adorns Porter's review and who was himself a Transylvanian in a broader sense.

ATTILA T. SZABÓ

Department of Biology,
Berzsenyi College,
9701 Sombathely,
Hungary

1. Porter, R. *Nature* **336**, 283 (1988).

2. Szabó, T.A. *Sr The Historical Thesaurus of Transylvanian-Hungarian Vocabulary Vols I-IV* (Kriterion, Bucharest, 1975-1984).

What prospects for *perestroika*?

The Soviet scientific community seems to have won a substantial measure of freedom in the past year, but the underlying problems of the Soviet economy cast a cloud over change.

Moscow

NOT even the unseasonably warm weather (on at least two days last week, the snow was melting) can explain people's cheerfulness — indeed, most would prefer that the weather would be what it should be at this time of the year, crisp and cold. Much of the explanation is that things have indeed changed a great deal, and for the better. Most noticeably, people who a year ago had never been abroad now seem to be regular travellers to what is still called the West (but which includes Japan). For those not travelling, it is now possible to find periodicals such as the *International Herald Tribune* in the hotels, but shockingly out of date.

There have been notable changes in the regulations affecting scientific work, and in the conscientiousness with which they are carried through, that have simplified once-complicated tasks. On travel, for example, an institute director's decision is usually, now, sufficient.

Similarly, with the publication of research reports, the formal approval of manuscripts by the parent academy (always a formality, but a thief of time) has been abandoned. An institute's director can approve publication in a Soviet journal, but the state copyright board which formally exploits overseas the intellectual property of Soviet citizens must still approve of publication in foreign journals. The decision of at least one writer to defy the board's authority (on the grounds that he would earn more without its intervention) has even raised the question of how long that system can continue to apply to scientific articles (where, usually, no money changes hands).

This betokens a kind of permissiveness that is unfamiliar. While the limits of *glasnost* remain to be explored fully (by experiment) and defined (presumably by reproof of some kind), even those in authority who make no secret of their uneasiness at present trends are evidently diffident at making too much fuss. But Mr Mikhail Gorbachev's 2-hour speech to a gathering of intellectuals at the Kremlin on 7 January, which argued among other things for support and also for self-restraint, may suggest that definition may not be far away.

Meanwhile, people have more elections to occupy their attention. In at least some of the Soviet Academy's institutes, this will be the third occasion in a year in which

researchers have been voting on each other's suitability for one or another function. Less than a year ago, institutes were nominating members of the special Party Conference, in October many were electing new directors, now they are nominating people to be deputies in the new constitution's scheme for a representative parliamentary system.

The idea is that the Soviet Union as a whole should elect some 5,000 people to a chamber of deputies that will, this summer, elect from its own number a body known as the Supreme Soviet (at present nominated and largely honorific) to consist of full-time scrutineers of government legislation. The Academy of Sciences has 25 seats in the electoral college (the Soviet Union's philatelists' society also has a quota), the Communist Party as such has 100, but other Party members will no doubt be elected as district representatives.

Academics Andrei Sakharov and Roald Z. Sagdeev, in that order, seem to be well ahead among the nominations of individual institutes. It is generally believed that, if elected, each would be prepared to stand for the Supreme Soviet. The academy's candidates will be chosen at a meeting this week. There is some ribaldry that one institute, having nominated its own director, decided not to nominate Sakharov as well for fear of undermining its own candidate's chances.

Meanwhile, it is certain that science will be represented at a senior level in the electoral college. Academician G.I. Marchuk, president of the academy, Academician V. Kotelnikov and Ye. Velikhov (both vice-presidents, the latter now also director of the Kurchatov Institute) together with two other members of the academy appear on the Communist Party's final slate of 100.

Sakharov himself seems gracefully to have made the transition from exile to statesman foreshadowed by his election to the praesidium (governing council) of the academy three months ago. Back from a first visit to the United States only in December, he has been, at the request of Mr Gorbachev, on a fact-finding visit to Nagorno-Karabakh (the Armenian enclave in Azerbaijan) and has caught the attention of the academy with a powerful speech on ecology. If there were still a vacancy as honorific head of state (Mr Gorbachev is now both president and Party general-secretary), Sakharov would

surely be appointed instantly.

Not all elections have brought such comfort. There is, for example, the sad case of Academician G. N. Basov, awarded a Nobel Prize in 1964 for the development of lasers (jointly with Charles W. Townes of the United States and his own Moscow colleague A.M. Prochorov) and for many years head of the Lebedev Institute in Moscow.

Basov's star began to set when he was not elected as the institute's nominee to the Party conference. Now he has been succeeded at a reorganized institute (strictly, a federation of research groups) by Academician Leonid Keldysh (son of a previous academy president). The final humiliation is that Basov has not been nominated as a deputy by the national "science-awareness" organization (analogous to the British Association for the Advancement of Science) of which he is the president. Yet he is a vigorous and even kindly man, but lacks adroitness.

Is it possible that, unused to free elections, the Soviet Union equates success in particular elections as a proof of all-round competence (and vice versa)? For that matter, it remains to be demonstrated that the election of institute directors by the staff whom they will manage will always produce satisfactory results.

Elections apart, people are worrying about *perestroika*, the process meant to put the Soviet Union back on its feet. Some welcome the fact that *perestroika* has further emptied the shops of goods on the paradoxical grounds that complacency is the real enemy, but the lack of a clear economic strategy is daunting.

The abrupt change two weeks ago of the rules affecting cooperatives (a kind of euphemism for private businesses) is a more particular worry. Cooperatives making medical diagnostics (small-scale biotechnology companies) will in future be regulated by ministries, others (as in publishing) are forbidden. The origins of this change are obscure, but seem to embody general discontent with the high prices charged by the cooperators, and the consequent high profits.

Meanwhile, too many rubles chase too few goods, so much so that offers to exchange rubles for foreign currency have become almost indefinitely open. The chief danger for *perestroika* in these circumstances is that of instability. Stop-gap foreign credits seem the only quick way out.

John Maddox

Protein structure

Polyhedra for helical proteins

Cyrus Chothia

THE principles that govern how α -helices assemble to form the structure of globular proteins have not previously been understood. But now, in a paper in the *Journal of Molecular Biology*¹, Murzin and Finkelstein present an elegant solution to the problem. Some time ago, these authors showed that the geometries of α -helices packed around a ball-like core can be described by quasi-spherical polyhedra. They went on to argue that the intrinsic properties of α -helices and of their packing surfaces would mean that all or most helical proteins would have a structure described by one of a small number of such polyhedra². In their new work¹, they demonstrate that this is indeed the case.

Three general features of the α -helices found in globular proteins are used in considering ideal helix packings and the polyhedra that describe the geometry of these packings. First, helices have the general shape of cylinders; second, they pack together with one face on the interior of the protein and the other exposed to the solvent; and third, they pack around a central hydrophobic core whose diameter

must be as spherical as possible and that the helices are close packed with a similar number of contacts. The polyhedra whose geometries reflect these properties are those that satisfy four criteria: a convex surface; all faces triangular; all ribs the same length; and a minimal difference in the number of ribs coming to each vertex. Murzin and Finkelstein find that only one polyhedron can satisfy these conditions for a given number of helices if the number does not exceed six. For more than six helices there are no such polyhedra. Ideal three-helix packings are described by an octahedron, four-helix packings by a dodecahedron, five-helix packings by a hexadecahedron and six-helix packings by an icosahedron (see Fig. 1c).

Helices can be placed on the ribs of polyhedra in several different ways. Three helices fit on the octahedron in two different arrangements: that seen in Fig. 1b and its enantiomorph. Four helices can be placed on the dodecahedron, and five helices on the hexadecahedron, in ten different ways. Six helices can be placed on the icosahedron in eight different ways.

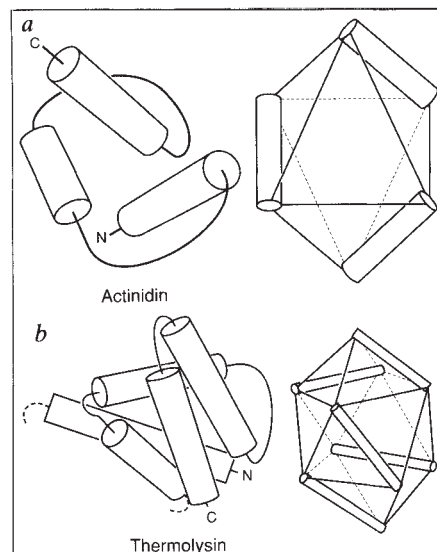


Fig. 2 Observed helix packing and ideal models. *a*, The three-helix packing found in actinidin³ and *b*, the five-helix packing found in thermolysin⁴. The drawings were made by fitting cylinders to the atomic coordinates of the helices. The ideal helix packings whose geometry corresponds to that observed are shown on the right.

three- and five-helix packings observed in actinidin³ and thermolysin⁴ and the corresponding ideal models.

The positions and orientations of helices in the observed structures differ from those found for helices in the corresponding ideal model by 2 Å and 20° on average. The larger deviations occur in proteins where the spherical nature of the packing is distorted by the helices clustering not about a point but about a somewhat more elongated or flattened region.

The absence of ideal polyhedra for assemblies of more than six helices suggests that in the few proteins that contain such assemblies the packing is described by different types of models. One possible alternative would be the formation of layer structures by the helices⁵, like those in proteins containing β -sheets^{6,7}. The low-resolution structure of bacteriorhodopsin⁸, for example, suggests that it is formed by seven helices packed in two layers.

Analyses of protein structures show that pairs of close-packed helices usually have their axes tilted in one of two directions: at angles of approximately -50° or +20°, arising from the ridge and groove character of helix surfaces⁹. The polyhedron models in which pairs of neighbouring helices are in orientations close to these values are those that most commonly correspond to the observed structures. There are two models for the packing of three helices, for example. That shown in Fig. 2a has the axes of the packed helices inclined at -60°, a value close to the preferred orientation of -50°, whereas in the alternative model the angle is +60°, which is some way from the nearest preferred value of +20°. Of the eight

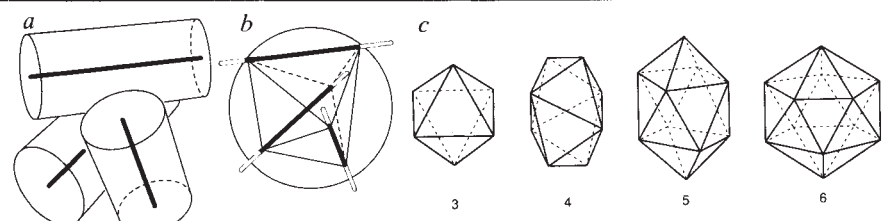


Fig. 1 Use of polyhedra to describe the geometry of helix packings. *a*, Three α -helices packed together are shown as cylinders with a diameter of 10 Å and with their axes as thick lines. *b*, The construction of the polyhedron that describes the geometry of this packing. A sphere is drawn from the centre of the packing with a radius of 11 Å. The parts of the helix axes enclosed within the sphere form one set of ribs of the polyhedron. The other set of ribs is formed by the lines joining the nearest ends of the helix axes. *c*, The quasi-spherical polyhedra for the ideal packing of three, four, five and six helices.

is the length of two residues. Three helices packed in this manner are shown in Fig. 1a.

Figure 1b shows how the geometry of this packing can be described in terms of a polyhedron. To construct this polyhedron, a sphere of radius 11 Å is first drawn from the centre of the packing. This sphere encloses the central section of each helix axis, and these sections form one set of the ribs of a polyhedron. The polyhedron is completed by another set of ribs formed by the connections linking the ends of the nearest helices (Fig. 1b). For the three-helix packing shown in Fig. 1a, the polyhedron constructed is the octahedron shown in Fig. 1b. Using this construction, the geometry of any packing of helices can be described by a polyhedron.

Murzin and Finkelstein argue that in real proteins the assemblies of helices

(Note in these arrangements the direction of the polypeptide chain is ignored: a helix is just treated as a cylinder.)

This 'quasi-spherical' polyhedron model describes the ideal packing of simplified approximations of α -helices. To determine if the model is relevant to real proteins, Murzin and Finkelstein compare the geometries of the helix arrangements observed in real protein structures with those expected from ideal models¹. They make comparisons for all the proteins that contain helix packings and for which atomic coordinates have been published. In almost all cases there is a close fit between the helix packings found in these proteins and one of those in the ideal models. The larger proteins are formed by a combination of simple polyhedral packings. I show in Fig. 2 the

non-homologous three-helix packings observed, seven are described by the first model and one by the second¹.

The quasi-spherical polyhedron model for the packing of helices has important implications for fields outside protein folding. Similarities in the structures of proteins are often taken to imply evolutionary relations even if there is no significant homology in sequence. Such a relation may exist. But Murzin and Finkelstein show that helical proteins can have very similar architectures independent of both sequence homology and chain pathway. Helix packings from non-homologous proteins but whose geometry is described by the same model appear very similar. Indeed, helix packings whose connectivities are very different but which fit the same model have sets of alpha

carbon atoms that superpose with a root-mean-square difference in position of less than 2.5 Å. Values somewhat greater than this have previously been taken as clear evidence for evolutionary relationships. □

1. Murzin, A.G. & Finkelstein, A.V. *J. molec. Biol.* **204**, 749–770 (1988).
2. Murzin, A.G. & Finkelstein, A.V. *Biofizika* **28**, 905–911 (1983).
3. Baker, E.N. *J. molec. Biol.* **141**, 441–484 (1980).
4. Holmes, M.A. & Matthews, B.W. *J. molec. Biol.* **160**, 623–639 (1982).
5. Efimov, A.V. *J. molec. Biol.* **134**, 23–40 (1979).
6. Chothia, C. *Rev. Biochem.* **53**, 537–572 (1984).
7. Finkelstein, A.V. & Ptitsyn, O. *Prog. Biophys. molec. Biol.* **50**, 171–190 (1987).
8. Henderson, R. & Unwin, P.N.T. *Nature* **257**, 28–32 (1975).
9. Chothia, C. *et al. J. molec. Biol.* **145**, 215–250 (1981).

Cyrus Chothia is in the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH and the Department of Chemistry, University College London, 20 Gordon Street, London WC1, UK.

Computation

Breaking the matrix speed limit

Steve McCormick

MATRICES abound in the sciences. Many problems involving inter-related variables can be cast into equations that use the familiar array form of a matrix. For large-scale problems involving many variables, the matrix in question is often composed mostly of zeros — a so-called sparse matrix. To capitalize on this structure, the equations can usually be solved by iteration on a computer. Unfortunately, the asymptotic approach of the iteration to the solution can become extremely slow, a process termed critical slowing down. Among the many possibilities for combating this stumbling block, that now used by R. G. Edwards *et al.* (*Phys. Rev. Lett.* **61**, 1333–1335; 1988) in a model problem shows great promise.

Sparse matrices arise in the modelling of large, complex phenomena from virtually every corner of science, including fluid mechanics, structural analysis, molecular dynamics, geodesy, economics, image reconstruction and X-ray crystallography. A sparse matrix expresses known local rules that govern the behaviour of a system of objects describing the phenomenon of interest. These rules characterize direct interactions that emerge as numbers in the matrix array. The numbers in a given row quantify the interaction of a given object with all of its neighbours. The location of the relatively few non-zero numbers in the row indicate who the neighbours are, and their magnitudes indicate the degree of the interaction. The actual behaviour of the system can be recovered numerically by solving a set of equations defined by the matrix, but conventional methods for solving these equations often suffer from the debility of critical slowing down.

In their new work, Edwards *et al.* investigate this sort of trouble in the context of certain lattice systems that arise in the study of random-resistor networks, quantum chromodynamics and discrete Schrödinger operators. (For example, random-resistor networks are a paradigm for the conduction of current through composite materials. The non-zero matrix elements correspond to resistors randomly placed between neighbouring points of a lattice, and the idea is to determine the current at these points for a given external current supplied to the lattice.) The cure Edwards *et al.* use involves a relatively new method, called algebraic multigrid (AMG), the basic idea of which can be described as follows.

Conventional methods start with a given approximation to the state of the system, then attempt to make improvements by a sequence of sweeps that adjust the state of each object in the system, forcing the objects in their turn to obey the local rules. Critical slowing down occurs after a few such sweeps because these local adjustments can take an inordinate amount of time to determine the global state of the system. This is not unlike the trouble that an ant has in building or, better, rebuilding an ant hill. AMG avoids this trouble by making additional improvements on more global scales.

This is accomplished systematically by aggregating the objects into groups, automatically translating the local rules to apply to the aggregated system, then sweeping through these groups to make adjustments at the aggregated level. This approach gives an ant the power to work with clusters of sand for adjusting the coarse features of the hill, reserving the

delicate manipulation of individual grains for the finer features. It gives the researcher the ability to determine the state of a system at a cost equivalent to just a couple of sweeps of the conventional method.

This approach is often relatively easy to develop, at least in principle. For lattice systems, there is usually one property (or variable) used to express the local rules among the objects, and this property can be treated as sort of a metric to determine appropriately how the objects and local rules should be aggregated. For example, in random-resistor networks, the resistance between two sites that are geometric neighbours can be interpreted as a measure of the intervening distance. Sites that are close in this sense can be grouped, and the aggregated rules can be determined by an averaging of the individual rules based on this measure. AMG computer software that automatically implements this approach has indeed been fairly successful for lattice systems, elliptic fluid flows and various other applications.

But for other complicated phenomena, development of the AMG approach can require very substantial research. For example, many systems use two or more properties per object to define local rules. In structural analysis, displacements for each coordinate and rotations for each pair of coordinates determine the state of the system. In developing an AMG scheme for a new problem of this type, just how such properties should be used to form distance measures is not always clear at the outset. But there have been very successful designs of AMG techniques for a broadening class of applications, including least-squares problems in geodesy, general fluid flows and various structural-analysis problems (see *SIAM Frontiers in Applied Mathematics* Vol. 3, ed. McCormick, S. F.; SIAM, Pennsylvania, 1987).

But this is just the beginning. Led by its pioneer A. Brandt and others, the multigrid concept has been making dramatic inroads into many new areas of science (see *Lecture Notes in Pure and Applied Mathematics* Vol. 110, ed. McCormick, S. F.; Springer, New York, 1988). Several challenging problems that were previously thought to be intractable now seem ready to fall before the multigrid axe. In the field of quantum chromodynamics, it is estimated that conventional methods could take ten million years or more to determine the mass of a proton. A novel multigrid approach seems to hold the promise of reducing this time to perhaps as little as a few hours. If such dramatic improvements can be made in this and other challenging fields, then critical slowing down may be well on its way out of scientific computation. □

Steve McCormick is in the Computational Math Group and the Center for Applied Parallel Processing of the University of Colorado, Denver, Colorado 80204, USA.

Gene transcription

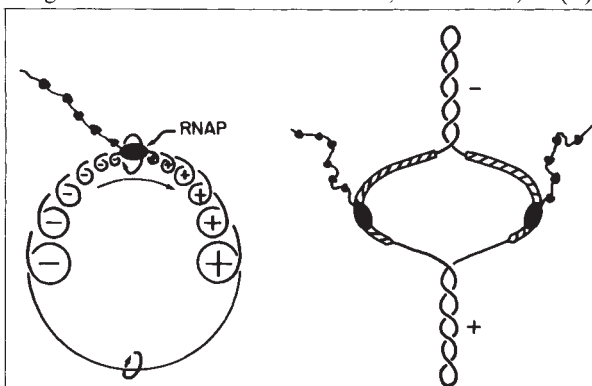
Waves of DNA supercoiling

Maxim Frank-Kamenetskii

IF YOU walk to the window and rotate the double string used to draw the blind in a clockwise direction long enough to obtain a double helix, then insert a pen between the two strands and push it ahead without rotation, you would be making a mechanistic model of the transcription process: the pen, or RNA polymerase, moves along DNA, the interwound double string. From this experiment, it can be seen that while RNA polymerase translocates it should overwind DNA in front of itself and underwind DNA behind itself. In other words, DNA becomes positively supercoiled in front of RNA polymerase and negatively supercoiled behind it (see figure). Though an inevitable consequence of the DNA helical structure, this wave of supercoiling which moves with RNA polymerase seems fantastic. However, the elegant recent experiments by J. C. Wang and co-workers demonstrate that these waves of supercoiling actually exist both in prokaryotic and eukaryotic cells (Wu, H.-Y. *et al. Cell* 53, 433-440; 1988; Giaever, G.N. & Wang, J.C. *Cell* 55, 849-856; 1988). These data will surely change the general attitude to the biological significance of DNA supercoiling and to the importance of DNA topology.

Continuing the experiment with the string and pen, very soon the pen stops because the double string cannot be overwound any more. Thus, one has to suppose either that DNA and RNA polymerase rotate around each other, or that the cell can eliminate both positive and negative supercoils. One can hardly expect that a very long DNA molecule and a very bulky transcription machinery loaded in pro-

karyotes with an even bulkier translation machinery could rotate around each other. On the other hand, topoisomerases are known to be able to change DNA supercoiling. On the basis of these simple reasons, L. F. Liu and Wang put forward the concept of waves of supercoiling (*Proc. natn. Acad. Sci. U.S.A.* 84, 7024-7027;



Formation of twin supercoiling domains during transcription. Left, a single RNA polymerase (RNAP) transcribing along a plasmid. Translocation generates a positively supercoiled domain (+) in front of RNAP and a negatively supercoiled domain (-) behind it. Right, when two opposing transcripts are present on the same circular DNA, the two supercoiled domains cannot merge by rotating the DNA alone; one of the transcription ensembles must be rotated as well. (From Wu, H.-Y. *et al. Cell* 53, 433; 1988.)

1987). But how can the wave of supercoiling be measured? Extraction of DNA from the cell, ridding it of proteins, causes the memory of the wave to be lost, as the wave of supercoiling does not change the overall DNA linking number. Although Wang and co-workers could not directly observe the wave of supercoiling *in vivo*, they unambiguously demonstrated its existence in experiments involving the inhibition of different DNA topoisomerases. Their most striking piece of

evidence is the formation of positively supercoiled plasmid DNA in *Escherichia coli* when DNA gyrase is inhibited. In this case, topoisomerase I continues to remove negative supercoils, while positive supercoils, which are normally removed by DNA gyrase, are accumulated in DNA.

In contrast with the prokaryotic case, one cannot selectively cut off the topoisomerase activities which relax positive and negative supercoils in eukaryotes. Giaever and Wang overcame this difficulty by using yeast temperature-sensitive (*ts*) strains transformed by a plasmid expressing *E. coli* DNA topoisomerase I. This enzyme relaxes negative supercoils but cannot relax positive ones. In full agreement with the model, an endogenous plasmid extracted from cells grown under nonpermissive temperature when yeast endogenous topoisomerase was inactive, turns out to be significantly positively supercoiled.

These findings throw a totally new light on the biological significance of DNA supercoiling. Specifically, people had thought that DNA gyrase in *E. coli* serves as an enzyme which introduces negative supercoils into DNA. Competing with topoisomerase I, it supports the native level of DNA supercoiling in the cell. Some held that by changing DNA supercoiling, gene expression could be regulated.

Now it can be seen that this picture, which was almost generally accepted, is really upside down. Indeed, DNA gyrase seems to eliminate positive supercoils rather than to create negative ones in *E. coli*. Native supercoiling is an irrelevant notion because actual local supercoiling may be highly positive, highly negative or negligible depending on the position of promoters, on the current position of RNA polymerase and on the relationship between the rate of RNA polymerase translocation and the efficiency of supercoil removal by the topoisomerases. Supercoiling depends on transcription to a far greater extent than transcription depends on supercoiling.

The findings of Wang and his colleagues emphasize the significance of the unusual structures in DNA (open regions, cruciforms, Z and H forms and so on) which form *in vitro* under negative supercoiling. These structures can form transiently in the negative domain of the wave of supercoiling (behind a moving RNA polymerase) and be trapped by specific proteins. Among other things, this could explain the transcriptional activation of replication and recombination. □

Maxim Frank-Kamenetskii is in the Institute of Molecular Genetics, USSR Academy of Sciences, Moscow 123182, USSR.

100 years ago

THE BALD-HEADED CHIMPANZEE



THERE is no longer any room for doubt amongst naturalists as to the complete distinctness of the larger anthropoid ape of tropical Africa, the gorilla, from its smaller brother, the chimpanzee. But, on the question whether there is only one chimpanzee over the African continent, there is still much difference of opinion. The acquisition of the fine female specimen, known by the name of "Sally", by the Zoological Society in 1883, caused me to change my views. There can be no doubt that this animal, when compared with specimens of the ordinary chimpanzee, presents very essential points of distinction. The uniform black face and nearly naked forehead, render "Sally" conspicuously different from the many specimens of the common chimpanzee that the Society has previously received.

From *Nature* 39, 254; 10 January 1889.

Geochronology

Diamond dating anomalies

Martin H. Dodson

DIAMONDS are particularly interesting samples of the Earth's mantle. Because of their unique combination of chemical stability and mechanical strength, impurities which they incorporate at their birth are protected almost indefinitely against interactions with the surrounding mantle. Some diamonds are known to be 2–3 thousand million years (Gyr) old^{1,2}, but direct dating of diamonds by the potassium–argon method has not been attempted until fairly recently^{3,4}, despite the great sensitivity of modern equipment for argon isotope analysis. Expectations, however, in geochronology are easily confounded; Ozima and co-workers, in a series of papers^{4–6} culminating in the one on page 226 of this issue⁶, show that forming diamonds can incorporate large amounts of 'excess' radiogenic argon in association with potassium, so that what began as a dating project has ended as a study of ancient mantle fluids.

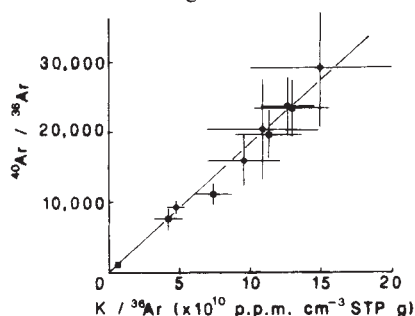
Because of previous indications that excess argon would be a problem³, Ozima and his colleagues decided to use an 'isochron' method (see figure) to date a suite of ten samples from Zaire called 'cubic' diamonds. These curious stones are polycrystalline aggregates with a rather fibrous structure and roughly cubic in shape, some having clear octahedral cores. It seemed likely that they are of uniform age and share the same initial argon isotope composition, essential conditions for the isochron method. In the event the stones yielded⁴ a respectable linear array on the potassium–argon isochron diagram — indicative of uniform age — but a thoroughly unrespectable age of 6 Gyr, apparently making them 1.5 Gyr older than the Solar System.

Obtained from a single sample, this result would simply have been rejected as being due to excess radiogenic argon. The linearity of the isochron plot, however, made such an idea hard to accept, for it required that the excess argon was incorporated in exact proportion to the potassium. This seemed unlikely, because normally the contrast in chemistry between potassium and argon ensures that the Ar/K ratio in a new mineral grain is both lower (commonly zero) and more variable than in its surroundings.

A more fundamental explanation was therefore sought, and it was suggested⁴ that the potassium had an anomalously high abundance of the parent ⁴⁰K. However, it was soon shown⁵ that potassium in Zaire diamonds is isotopically indistinguishable from that found in any other terrestrial material. Thus the diamonds

must either have incorporated a pre-solar component (diamonds have been found in meteorites) or have been isotopically contaminated in a very unusual way.

The question is now resolved in this issue⁶ in favour of contamination during crystal growth, by mechanical entrapment of an argon-rich fluid. The clinching evidence comes from the argon–argon (or argon-39) method, in which samples are irradiated with fast neutrons, converting a small proportion of ³⁹K to ³⁹Ar. The ratio ⁴⁰Ar/³⁹Ar in gas extracted from the



Isochron plot for cubic diamonds from the earlier paper by Ozima and co-workers⁴. The ⁴⁰Ar and K contents of samples are plotted as ratios to their ³⁶Ar contents, which are assumed to be constant through time. Because ⁴⁰K decays radioactively to ⁴⁰Ar, data from samples formed at the same time, with uniform 'initial' argon-isotope composition (⁴⁰Ar/³⁶Ar), but variable K/³⁶Ar, would all fall on a straight line whose slope gives the age of the samples. The intercept shows the initial ⁴⁰Ar/³⁶Ar, in this case close to zero (-700 ± 900), suggesting that no ⁴⁰Ar was trapped in the forming diamonds. The ³⁹Ar data of Ozima and co-workers in this issue, however, show that these diamonds entrapped, as they grew, a variable mixture of fluids, one with high, fairly uniform Ar/K, another with low ⁴⁰Ar/³⁶Ar.

samples corresponds to the radiological daughter–parent ratio ⁴⁰Ar/⁴⁰K in the conventional method. The gas is usually extracted in stages while the sample is heated to successively higher temperatures; changes in apparent age during this step-heating procedure show differences between the locations of ⁴⁰Ar and ³⁹Ar within the sample. (Furthermore, during neutron irradiation ³⁸Ar is produced from chlorine and ³⁷Ar from calcium, so the concentrations of these elements may also be determined by argon isotope analyses.)

In their new paper, Ozima *et al.*⁶ present step-heating analyses of four irradiated diamond samples. The 'age' from their isochron plot is again close to 6 Gyr, with little indication of changes during step-heating. More important, they find an excellent correlation of chlorine with ⁴⁰Ar. Ozima *et al.* point out that similar

correlations obtained from fluid inclusions in the upper crust reflect leaching of argon and chlorine from crustal rocks. Evidently these diamonds provide evidence for analogous processes in the mantle, and the gases extracted from them consist of mixtures, in varying proportions, of a chlorine-free component with ⁴⁰Ar/³⁶Ar of about 340 and a component with ⁴⁰Ar/Cl of 8 or 9. Rather less convincing is the attempt, by means of a correlation diagram for radiogenic ⁴⁰Ar/K against Cl/K, to separate the ⁴⁰Ar into an excess component and an *in situ* radiogenic component (with a calculated age in the range 0–3 Gyr).

Navon *et al.*⁷ (see also ref. 8) recently presented elemental analyses for Zaire cubic diamonds rich in sub-micrometre inclusions, interpreting them in terms of volatile-rich fluids roughly comparable with potassic magmas. Ozima *et al.* give potassium, chlorine and calcium abundances which are broadly consistent with those of Navon *et al.*, and provide complementary argon-isotope data. The significance of these inclusions for the understanding of mantle processes⁸ would be hard to exaggerate, and it is clear that there must be further attempts to date them. Reported success in dating of fluid inclusions in crustal rocks⁹ suggests that the Rb–Sr method could be worth a try.

Finally, do the problems encountered in dating of diamonds have serious implications for more conventional applications of the potassium–argon method? The answer is, probably not. In these diamonds the potassium–argon system is evidently dominated by a rather homogeneous ⁴⁰Ar-rich fluid which has been incorporated mechanically during crystal growth, mixed with small amounts of a low ⁴⁰Ar/³⁶Ar component. Thus the system does not satisfy one of the most fundamental requirements for successful dating, that the radiometric clock must be reset by a process of chemical differentiation between parent and daughter elements. In ordinary rocks and minerals the immense difference between the chemical properties of potassium and argon ensures that this condition can usually be satisfactorily met, with the likelihood of obvious inconsistencies in the data if it is not. □

1. Kramers, J.D. *Earth planet. Sci. Lett.* **42**, 58–70 (1979).
2. Richardson, S.H., Gurney, J.J., Erlank, A.J. & Harris, J.W. *Nature* **310**, 198–202 (1984).
3. Ozima, M., Zashu, S. & Nitoh, O. *Geochim. cosmochim. Acta* **47**, 2217–2224 (1983).
4. Zashu, S., Ozima, M. & Nitoh, O. *Nature* **323**, 710–712 (1986).
5. Podosek, F.A. *et al. Nature* **334**, 607–609 (1988).
6. Ozima, M., Zashu, S., Takigami, Y. & Turner, G. *Nature* **337**, 226–229 (1989).
7. Navon, O., Hutcheon, I.D., Rossman, G.R. & Wasserburg, G.J. *Nature* **335**, 784–789 (1988).
8. Menzies, M.A. *Nature* **335**, 769–770 (1988).
9. Darbyshire, D.P.F. & Shepherd, T.J. *J. geol. Soc. Lond.* **142**, 1159–1177 (1985).

Martin H. Dodson is in the Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK.

Vertebrate neurobiology

Sound localization mechanisms

David Moore

DIFFERENCES in the intensity and time of arrival of a sound at the two ears are well known to be the primary cues for auditory spatial localization in vertebrates (a in the figure). But despite many speculations, the neural mechanisms underlying these processes are not well understood. In two new reports^{1,2}, Masakazu Konishi and colleagues demonstrate the cellular and topographical neural organization underlying the first stages of processing of these signals in barn owls.

Barn owls are one of a small number of owl species to have evolved bilateral ear asymmetry. Such species are suited for studies of sound-localization mechanisms,

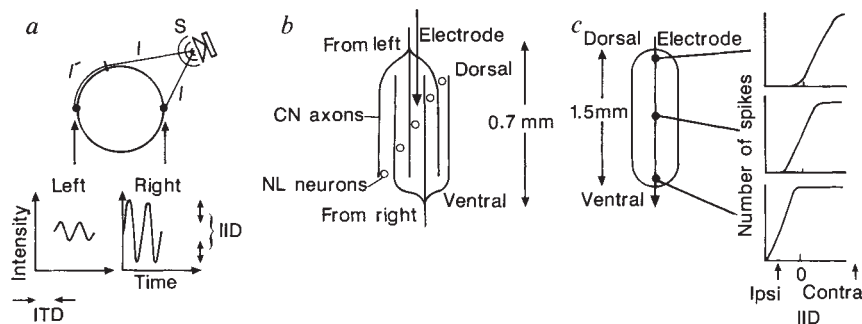
would travel through longer axons to arrive at left nucleus laminaris neurons coincidentally with signals that had been generated later in the left cochlear nucleus. This model also predicts that nucleus laminaris neurons should be topographically arranged so that coincidence detection corresponding to different regions of space varies systematically across the extent of the nucleus.

Carr and Konishi now provide¹ direct evidence for both these proposals. By filling single axons from the cochlear nucleus with the tracer horseradish peroxidase, they show that axons derived from the two ears enter the nucleus

acoustic stimulation of one ear are inhibited by stimulation of the other ear. The final level of the response is determined by the strength of excitatory and inhibitory input to each neuron. This, in turn, is determined partly by the intensity of the stimulus delivered to each ear and partly by factors intrinsic to the brain. For some neurons, inhibition dominates and may be complete when sound of equal intensity is presented to the two ears. For other neurons, sound presented to the inhibitory ear may need to be 20–30 decibels more intense than sound presented to the excitatory ear before any response suppression is observed.

In the ventral nucleus of the lateral lemniscus of the barn owl, Manley, Köppl and Konishi² show that neurons are organized topographically according to the IID for which their response is equally influenced by the excitatory and inhibitory ears (c in the figure). In the dorsal

a, The physical cues for sound localization are established by a difference in the path length (l') for sound reaching each ear, resulting in an ITD, and by the sound-shadowing effect of the head and outer ear producing an IID. Two cycles of a tone stimulus produced at the sound source (S) are shown as they would be recorded by microphones placed in the position of the ears. b, Schematic representation of axons from the left and right cochlear nucleus (CN) interdigitating within the barn owl nucleus laminaris (NL). Physical ITDs are offset by the time taken for spikes to travel through the depth of the NL from each side. According to their depth within the nucleus, the NL neurons show receive coincident spikes from the input axons at different ITDs. Arrow, direction of electrode penetrations made by Carr and Konishi¹. c, Electrode penetration (arrow) through the posterior part of the ventral nucleus of the lateral lemniscus (VLVp), showing (right) the responses to IID of neurons recorded at three depths. The IID at which balanced excitatory and inhibitory influences for each neuron was obtained is indicated. In dorsal VLVp, this IID is obtained when the contralateral ear intensity is higher (contra $\uparrow$). In ventral VLVp, the intensity at the ipsilateral ear needs to be higher (ipsi $\uparrow$) to give balanced excitation and inhibition (from ref. 2).



because interaural time difference (ITD) contributes exclusively to localization in the horizontal plane (azimuth)³ and interaural intensity difference (IID) forms the sole basis for vertical plane (elevation) localization⁴. The time and intensity pathways form separate streams in two divisions of the cochlear nucleus^{5,6}, the first synaptic station in the brain's auditory system. Neurons of the cochlear nucleus receive input from only one ear and so cannot detect ITD and IID. Detection therefore begins at the next level of the auditory system, where ITD processing occurs in the nucleus laminaris and IID processing in the posterior division of the ventral nucleus of the lateral lemniscus.

Jeffress proposed 40 years ago that ITD processing would be based on delay lines and coincidence detection⁷. He predicted that the length of axons derived from each ear and synapsing with binaurally innervated neurons should vary to compensate for physical time differences between the ears created by laterally positioned sounds. Thus, for a sound located on the right side of a barn owl's head, signals generated in the right cochlear nucleus

laminaris from opposite sides and interdigitate (b in the figure). Physiological recordings from these axons reveal progressive shifts in the timing (phase) of responses relative to tone stimuli delivered to each ear. Through the depth of the nucleus laminaris (0.7 millimetres), timing shifts for stimulation of either ear are about 200 microseconds, consistent with the behaviourally relevant range of ITDs. But timing shifts for the two ears are in opposite directions, as predicted from the anatomy. Recordings from the dorsal side of the nucleus laminaris show a time lead for the ear on the same side as the recording, whereas ventral recording sites show a time lead for the other ear. Thus, single axons of different peripheral origins and different sensitivities to sound frequency form an orderly representation of ITD through the depth of nucleus laminaris.

The mechanisms underlying IID processing are not as well understood, either theoretically or empirically. Nevertheless, it has been recognized for many years⁸ that there is neural sensitivity to IID. Most commonly, excitatory responses to

region this occurs for intensity differences where the input from the contralateral ear is more intense, and in the ventral region it occurs when the ipsilateral stimulus intensity is higher. Other systematic variations in response parameters across the dorso-ventral extent of the nucleus are correlated with this topography: excitatory and inhibitory response strengths and thresholds; spontaneous discharge levels; and sharpness of IID sensitivity. The results suggest that simple modulation of these parameters, particularly inhibitory response strength, can explain the observed topography.

Sound localization by humans and other

1. Carr, C. & Konishi, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8311–8315 (1988).
2. Manley, G., Köppl, C. & Konishi, M. *J. Neurosci.* **8**, 2665–2676 (1988).
3. Moiseff, A. & Konishi, M. *J. Neurosci.* **1**, 40–48 (1981).
4. Knudsen, E. I. & Konishi, M. *J. comp. Physiol.* **A133**, 13–21 (1979).
5. Takahashi, T., Moiseff, A. & Konishi, M. *J. Neurosci.* **4**, 1781–1786 (1984).
6. Sullivan, W. E. & Konishi, M. *J. Neurosci.* **4**, 1787–1799 (1984).
7. Jeffress, L. A. *J. comp. Physiol. Psychol.* **41**, 35–39 (1948).
8. Rose, J. E. *et al. J. Neurophysiol.* **29**, 288–314 (1966).
9. Moore, B. C. J. *An Introduction to the Psychology of Hearing*, 2nd edn (Academic, London, 1982).
10. Moore, D. R. *Br. med. Bull.* **43**, 856–870 (1987).

terrestrial mammals is almost certainly more complex than in the barn owl. Instead of acting over the same, relatively restricted frequency range to code for azimuth and elevation, time and intensity differences between the ears of mammals provide cues over different portions of the audible frequency range and both seem to be predominantly used for coding stimulus azimuth. Elevation cues in mammals are thought to result from sound-spectral transformations by the head and by the pinna, the visible part of the outer ear⁹. In fact, information derived from just one ear can, under certain stimulus conditions, lead to sound localization performance in humans that is as accurate as that achieved using two ears⁹.

Nevertheless, several key nuclei in the mammalian auditory pathway bear striking anatomical and physiological similarities to those in the barn owl known to be involved in localization. A topographic representation of IID has been demonstrated in the midbrain of the cat, for example, and two-dimensional neural 'maps' of auditory space have been found in homologous brain regions of several owl and mammalian species¹⁰. It therefore seems likely that at least some of the mechanisms for processing binaural sound localization cues are common in many vertebrate species. □

David Moore is in the Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK.

Ocean Drilling Program

Breakup of Gondwanaland

*Leg 122 shipboard scientific party**

OLD, sediment-starved, passive continental margins are important for elucidating the thermal and structural processes that lead to rifting and continental breakup. These regions also contain clues about global palaeoenvironmental and climatic changes, as well as the record of sea-level fluctuations. There have been few deep-sea drilling studies of the older parts of continental margins even though their unique record reaches back to the Jurassic and Triassic, when there was only a single continental mass — Pangaea. Pangaea was split subsequently by the Tethys Ocean into Gondwana in the south and Laurasia in the north. We have studied the old Gondwanan passive margin of Exmouth Plateau, off the coast of north-western Australia, to elucidate the evolution of rifting and drifting, in addition to the environmental changes that accompany and follow such momentous events.

The Exmouth Plateau is a rifted and deeply subsided fragment of continental crust, covered by more than 8 km of sediments, including nearly 1–2 km of post-breakup sediments. The present configuration of this region was initiated in the

Late Triassic to Early Cretaceous (230–130 million years ago, Myr) by East-Gondwanan rifting between northwestern Australia, greater India and other Gondwanan fragments to the north. Leg 122 data suggest that whereas the northern margin of the Exmouth Plateau bordering the Argo Abyssal Plain had rifted before the Late Berriasian (140 Myr), the western margin bordering the Gascoyne Abyssal Plain developed during the late Valanginian (135 Myr) when sea-floor spreading was initiated there.

We drilled at six sites (Fig. 1). Four of these (759–761 and 764) on the Wombat Plateau yielded a composite record of sediments deposited during rifting in the Triassic (248–213 Myr) and during the Cretaceous (144–65 Myr) and Cenozoic

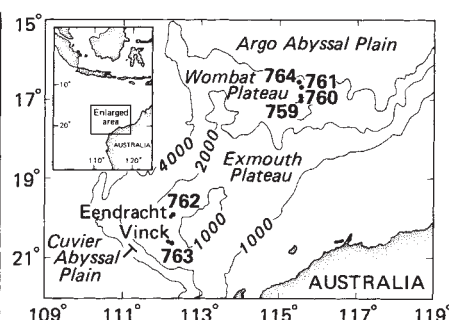


Fig. 1 Leg 122 sites (●) on the Exmouth Plateau. Commercial well sites are also indicated (•). Bathymetry given in corrected metres. (After von Stackelberg, U. *et al.* *BMR J. Aust. Geol. Geophys.* 5, 113; 1980.)

(65 Myr to the present) following the breakup. The other two sites (762 and 763) reveal the Cretaceous to Cenozoic record of palaeoenvironmental and passive margin evolution of the western part of the central Exmouth Plateau. This record consists of a thick clastic shelf-margin wedge that prograded from a southern source area during the Early Cretaceous late-rift phase, and was there overlain by an Upper Cretaceous to Cenozoic pelagic (marine) sequence.

Deep crustal extension and thinning of the northern Exmouth–Wombat Plateau area occurred during the Permian (286–248 Myr), and was followed by rifting during the Triassic. Carbonate rocks were deposited on the Wombat Plateau during the mid-Carnian (in the Triassic, around 228 Myr), in a southern embayment of a shallow Tethys Sea. The plateau periodically re-emerged until the Rhaetian (ending the Triassic, 213 Myr) when it developed a fully marine carbonate sequence. These carbonates resemble their coeval strata in the western Tethys region of the Alps.

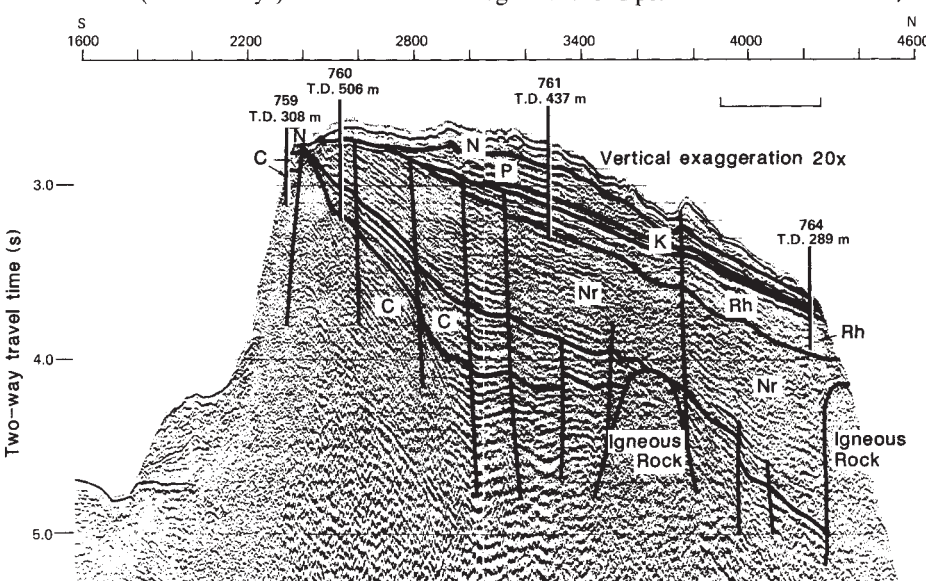


Fig. 2 Interpreted multi-channel seismic section of the Wombat Plateau showing the location of sites 759–761 and 764. Letters within the seismic profile refer to ages as follows: N, Neogene; P, Palaeogene; K, Cretaceous; Rh, Rhaetian; Nr, Norian; and C, Carnian. Scale bar, 10 km.

*Co-Chief Scientists: Ulrich von Rad (BGR, Hannover) and Bilal U. Haq (NSF). ODP Staff Scientist, Suzanne O'Connell (ODP, Texas A&M Univ.). Also, Alistair Bent (BP, UK), Charles D. Blome (U.S. Geol. Survey), Peter Borella (Saddleback College, California), Ronald Boyd (Dalhousie Univ.), Timothy J. Bralower (Florida Intl Univ.), Wolfram W. Brenner (Tübingen Univ.), Eric H. De Carlo (Hawaii Inst. Geophysics), Thierry Dumont (Institut Dolomieu, Grenoble), Neville Exon (Bureau of Mineral Resources, Canberra), Bruno Galbrun (Univ. Paris VI), Xenia Golovchenko (Lamont-Doherty Geol. Observ.), Naci Görür (Istanbul Tech. Univ.), Makoto Ito (Chiba Univ., Japan), Juan M. Lorenzo (Lamont-Doherty Geol. Observ.), Philip A. Meyers (Univ. Michigan), Ian Moxon (Stanford Univ.), David K. O'Brien (Hawaii Inst. Geophysics), Motoyoshi Oda (Kumamoto Univ.), Massimo Sarti (Univ. Calabria), William G. Siesser (Vanderbilt Univ.), Lloyd R. Snowdon (Geol. Survey Canada), Cheng Tang (Univ. California, Santa Cruz), Roy H. Wilkens (Hawaii Inst. Geophysics), Paul E. Williamson (Bureau of Mineral Resources, Canberra), Antonius A.H. Wonders (BP, UK).

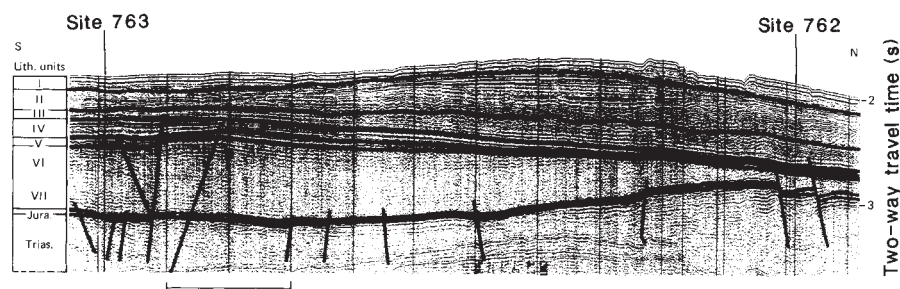


Fig. 3 Interpreted single-channel seismic line through sites 762 and 763 on the central Exmouth Plateau. Note thinning of the clastic wedge (units VI and VII) and thickening of the lower Tertiary sequence (unit II) from south to north. Scale bar, 7 nautical miles.

During the Jurassic (213–144 Myr), northward tilting raised Wombat Plateau above sea level again, and non-deposition or erosion prevented the preservation of Jurassic sequences before the onset of sea-floor spreading in the Argo Abyssal Plain. The plateau subsequently subsided rapidly during the early Cretaceous.

The time of breakup of the western and southern margins of the Exmouth Plateau is constrained by a major erosional unconformity, to around the end of the Neocomian (131–119 Myr), which agrees well with dating from sea-floor magnetic anomalies. Intervals of rising water depth are represented by condensed sections of thin glauconitic limestones and belemnite-rich mudstones. During the mid-Cretaceous (119–91 Myr), when the southern hinterland was drifting north-west with greater India, the southern supply of eroded rocks was cut off. Cyclic deposition of deep-water claystone during the middle to late Aptian (around 115 Myr) marks the onset of hemipelagic deposition (the 'juvenile ocean' stage).

A stagnant period of oxygen depletion

produced a well developed black shale with up to 15 per cent organic carbon at the Cenomanian/Turonian boundary (91 Myr). Post-Cenomanian deposition became dominated by pelagic carbonate deposition, continuous in the northern part of the basin (site 762) and reaching site 763 to the south in the middle Eocene (around 50 Myr).

After isolating unconformities caused by tectonic events, we used a sequence stratigraphic approach to decipher sea-level fluctuations at each site. Our preliminary results indicate that there are important sequence boundaries on the Wombat Plateau between sediments representing the middle and late Carnian, at the Norian/Rhaetian boundary and in sediments of latest Rhaetian age, whose timing conforms well with global changes. Sequence boundaries recognized in the Lower Cretaceous of the central Exmouth Plateau also correlate well with global changes but indicate that the pattern of accumulation of deposits may need to be modified in regions of fast sediment supply. □

Evolutionary biology

Seed of destruction

H. C. J. Godfray and Paul H. Harvey

WHEN Jack Werren and his colleagues attempted¹ to select for the sex ratio in *Nasonia vitripennis*, a parasitoid wasp that attacks houseflies and related species, they found to their surprise that females producing only sons started to appear in the line being selected for male-biased sex ratios. They established by crosses using genetically marked flies that the trait (*psr*, for paternal sex ratio) is extra-chromosomally inherited. This was the first recorded occurrence of cytoplasmic inheritance through the paternal line of an animal. In their most recent work^{2,3}, these authors establish the nature and mode of action of *psr*, and provide an explanation for its evolutionary origin. The serendipitous discovery of what has been termed the most selfish genetic element that has yet been described offers the prospect of new insights into the structure and function of

eukaryotic chromosomes.

Initially, it was uncertain whether *psr* was an infectious venereal disease which causes the death of female progeny, or a cytoplasmic factor transmitted to the egg during fertilization which causes it to become male. *N. vitripennis*, like other hymenopterans, has a haplodiploid genetic system with haploid males developing from unfertilized eggs and diploid females from fertilized eggs. As with many other hymenopterans, the female wasp produces a male-biased sex ratio when her sperm reserves are depleted, and a female-biased sex ratio in circumstances where her progeny tend to mate among themselves^{4,5} or when an appropriate maternally inherited cytoplasmic sex-ratio factor is present⁶. The proportion of progeny carrying the *psr* trait was highest when a female-biased sex ratio was

expected⁵. It therefore appeared that *psr* enters the putative female eggs at fertilization, converting the zygotes to males.

Cytological examination revealed that sperm carrying the *psr* trait enter and fertilize the egg as normal. During the first mitotic division, however, half the chromosomes appear clumped together in a dense mass of chromatin and do not participate in further cell divisions². The embryo, with its now haploid chromosome complement, develops as a normal male wasp. Genetic marker studies reveal that it is the paternal genome which is eliminated.

Extra chromosome

In their new work, Nur, Werren and co-workers reason that *psr* might be discovered by identifying the DNA present in infected but not in uninfected males. They created a λ -phage genomic library from *psr* male DNA, and searched for clones that hybridized with radioactively labelled whole DNA from *psr* males but not with DNA from normal males. They found 36 such clones, and subjected several of them to restriction-enzyme analysis. It turns out that much of the *psr*-specific DNA is in the form of satellite-like, short tandem repeats, the most abundant sequence being 171 base pairs long. This finding is not compatible with the idea that *psr* is a virus or a transposon, but suggests that it could, after all, be a supernumerary chromosome because satellite-like DNA is characteristic of heterochromatized regions of chromosomes. Indeed, detailed cytological study reveals a small chromosome found only in *psr* males.

Supernumerary, or B chromosomes are fairly common in animals and some even show varying degrees of meiotic drive⁷. But the devastating selfishness of *psr* is unique: in a mating between a male bearing *psr* and a female, *psr* first destroys the chromosomes of its original host and then, in the next generation, those of its newly adopted host. Though much of the DNA in the B chromosome is in the form of tandem repeats, specific regions are known to code for a *trans*-acting factor involved in the super-condensation of the parental chromosomes. Analysing the mode of action of *psr* could reveal the mechanism of chromosome condensation, which is poorly understood.

Where does the B chromosome come from and what is its fate? There are now preliminary answers to both questions⁸. There are several cytoplasmically incompatible strains of *N. vitripennis*, and in matings between wasps from two such strains, the paternal chromosomes are broken down. But it is possible to recover fragments of partially destroyed chromosomes that display high transmissibility through the male line but much lower

transmissibility through the female⁹. Such a fragment is under strong selective pressure to change the sex of the female zygotes it normally finds itself in to more congenial male zygotes.

Method of spread

Once *psr* has evolved, it may spread. Werren points out¹⁰ that if *psr* is transmitted by all the sperm of an infected male in a spatially unstructured population, the equilibrium frequency of *psr* is $(2x-1)/x$, where x is the proportion of eggs that female wasps fertilize. This means that *psr* will not persist unless $x > 0.5$, that is, unless female wasps fertilize more than half their eggs, which would mean producing a female-biased sex ratio in an uninfected population. *N. vitripennis* does indeed do this because its populations are spatially structured with mating occurring among the offspring of small groups of females, with consequent very high levels of local mate competition^{11,12}. Although this female-biased sex ratio favours the spread of *psr*, another feature of spatially structured populations tends to hinder its spread: if all the females laying their eggs in a single patch (perhaps a localized population of housefly larvae) have mated with *psr* males, all their progeny will be sons who will die without mating. Theoretical models of the spread of *psr* in spatially structured populations indicate that persistence is possible only for intermediate levels of population subdivision¹³.

Ironically, the persistence of *psr* may be markedly enhanced by the presence of maternally inherited cytoplasmic factors that cause female-biased sex ratios¹³, two of which are known from *N. vitripennis*^{6,14}. No systematic search for extrachromosomal sex-ratio factors has been undertaken in Hymenoptera other than in *N. vitripennis*. Of all groups with variable sex ratios, parasitic wasps were thought to be the best understood: perhaps there is a whole new level of complexity yet to be explored. □

1. Werren, J.H., Skinner, S.W. & Charnov, E.L. *Nature* **293**, 467–468 (1981).
2. Werren, J.H., Nur U. & Eickbush, D. *Nature* **327**, 75–76 (1987).
3. Nur, U., Werren, J.H., Eickbush, D.G., Burke, W.D. & Eickbush, D. *Science* **240**, 512–514 (1988).
4. Hamilton, W.D. *Science* **156**, 477–488 (1967).
5. Werren, J. & van den Assem, J. *Genetics* **114**, 217–233 (1986).
6. Skinner, S.W. *Science* **215**, 1133–1134 (1982).
7. Jones, R.N. & Rees, H. *B Chromosomes* (Academic, New York, 1982).
8. Werren, J.H. *Am. Nat.* (in the press).
9. Ryan, S.L., Saul, G.B. & Conner, G.W. *J. Hered.* **76**, 21 (1985).
10. Werren, J.H. *J. theor. Biol.* **124**, 317–334 (1987).
11. Werren, J.H. *Science* **208**, 1157–1159 (1981).
12. Werren, J.H. *Evolution* **37**, 116–124 (1983).
13. Werren, J.H. & Beukenboom, P. *J. theor. Biol.* (in the press).
14. Skinner, S.W. *Genetics* **109**, 745–759 (1985).

H. C. J. Godfray is in the Department of Pure and Applied Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY; Paul H. Harvey is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Raymond Arthur Dart (1893–1988)

RAYMOND Dart died in Sandton, Johannesburg, on 22 November 1988, almost 64 years to the day since he had received the chunk of sandy limestone containing the Taung skull that was to create his reputation. Dart recognized certain extraordinary features of the skull — its relatively erect poise, small canine teeth and endocranial capacity, scarcely differing from those of extant apes. But he noted also that the form of the brain was human-like in its parietal lobe. In particular, he recognized the impression of what he took to be the lunate sulcus. It was well back, as in modern man, and not well forward as in modern apes. He then made a great intellectual leap: he

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Dart — effected a revolution in knowledge. interpreted this curious blend of traits as evidence for the former presence in Africa of an anthropoid that had moved substantially in a human direction. This he claimed in his historic paper in *Nature* on 7 February 1925. He named the species *Australopithecus africanus*.

Many deterrents to the acceptance of Dart's claim were predicated on the accepted wisdom of the day: the idea that Asia, not Africa, had cradled mankind; and the notion that the brain had expanded early in hominization, a view supported by the features of the ill-famed Piltdown bones. Moreover, the Taung specimen was of a child and, it was averred, the nature of the adult could not have been predicted. Arthur Keith, with most later investigators, rejected the evidence of the lunate sulcus, holding that its position could not be identified with certainty on the Taung endocrast.

The main controversy Dart stirred up did not abate for 25 years, when many new specimens from the Transvaal were compared with the skeletons of modern apes. It was realized that, even if one discounts the lunate sulcus, enough other features mark the specimen not only as unique, but as morphologically intermediate in several respects between apes and men. Only then did it emerge that Dart's claim had been basically correct and, indeed, that there was every justification for classifying the fossils as members of the Hominidae.

Dart's recognition and interpretation of the significance of *Australopithecus*

remains one of the great steps forward in the history of palaeoanthropology. With hindsight, the essence of his 1925 contribution lay in his willingness to overlook the small brain size of *Australopithecus*, in the face of other strong pointers to its evolutionary status. Brain-size alone, he declared in a 1956 paper, was not the acid test; the form of the brain was more important. Late in life, Keith, the first of Dart's main opponents, generously acknowledged that he had been wrong and Dart right: he proposed to call the australopithecines 'Dartians'.

After World War II, following discoveries of fossil baboons at Makapansgat, northern Transvaal, Dart was lured back into the field — and a new chapter opened. Intrigued by the thousands of fossilized, broken antelope bones in that cave deposit, he proposed in 1955 that the bones had been broken deliberately, fashioned and used by the australopithecines. His proposed Osteodontokeratic culture was based on the use as implements of bones, teeth and horn-cores, especially of ungulates. This notion has now been largely rejected, as evidence has built up that big cats, hyenas and porcupines were probably the bone breakers and accumulators.

But Dart's hypothesis had a remarkable spin-off. When he first proposed his bone-tool theory in 1955, scarcely anything was known of how animals, or agents like sunlight and water, affected bones after death. To disprove Dart's proposal, a host of new studies were made of the effects of such biotic and physical agencies on bones. A new scientific discipline came into being, called taphonomy. Dart's ideas were a major catalyst in the birth of this new field. Thus, it can be claimed of Dart that he effected a revolution in knowledge about man's place in nature, and that indirectly he triggered the foundation of a new scientific discipline.

Raymond Dart was born during the flood that swept Queensland, Australia, in 1893. He read science at the new University of Queensland, Brisbane, and medicine at Sydney. After serving in the Australian Army Medical Corps, he worked under Elliot Smith in the Anatomy Department at University College, London. There followed six months as a Rockefeller fellow in the United States. In 1923, he became head of the Department of Anatomy at the newly formed University of the Witwatersrand Medical School, a position he held for 36 years. Dart was a prime architect of the Medical School and a maker of men. The Raymond Dart Collection of Human Skeletons is one of the tangible monuments to his initiative and foresight. Phillip V. Tobias

Phillip V. Tobias is Professor of Anatomy and Human Biology at the University of Witwatersrand Medical School, Parktown 2193, South Africa.

Active galactic nuclei

Pointing towards the source?

Virginia Trimble

QUASARS, like beauty, may be in the eye of the beholder. So suggests Peter D. Barthel in a paper in the latest *Astrophysical Journal*¹ that collates observations collected by his² and many other groups of astronomers. According to his version of the unified scheme for active galactic nuclei, radio-emitting quasistellar objects (QSOs) and powerful radio galaxies are the same sorts of objects, seen from different directions, and radio-quiet QSOs may be the similarly oriented subset of powerful galaxies with bright infrared sources (called IRAS galaxies after the satellite that measured their fluxes).

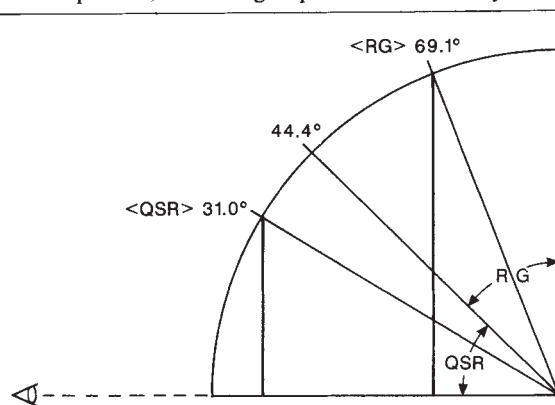
The generic name, active galactic nuclei (AGNs), indicates a long-standing recognition of some kind of unity within the zoo. The beasts on display include (a) quasars (an attempt at pronouncing the acronym QRS for quasi-stellar radio source), which are optically compact, show broad emission lines at large redshifts, and emit powerfully at radio wavelengths, (b) QSOs (much more numerous and not strong radio sources, but otherwise similar), (c) BL Lac objects (with rapid variability, strong polarization, and weak or no emission lines; named for their prototype), (d) radio galaxies, (e) optically violent variables (OVVs) and (f) Seyfert galaxies (bright nuclei and strong emission lines with moderate redshifts, named for their discoverer). The host galaxies are generally thought to resemble ellipticals for types (a) and (d), spirals for (b) and (f). Many are interacting galaxies.

All AGNs exhibit the astrophysical difficulty of producing lots of energy from small volumes with the capacity for very rapid change in both brightness and location of the emitting regions. And all are currently explained by a standard model^{3,4}. Its components are a massive black hole (10^5 – 10^9 times the mass of our Sun) at the centre of a galaxy; energy generation and transport by some combination of gas flowing inward and magnetic fields linking the hole to more distant gas; and one or two outgoing, well-collimated jets of particles, field and photons moving at speeds close enough to the speed of light that special relativity must be applied rather carefully to avoid inconsistencies in interpreting the observations⁵.

The model clearly provides a wealth of parameters that can be varied among assorted kinds of AGNs — the mass of the black hole, the rate of gas inflow (relative

to an upper limit set by radiation pressure), the field strength and configuration, the geometry of surrounding gas, the mechanism and precision of jet collimation, the dominant components of the jet (plasma of electrons and positrons or electrons and protons; magnetic field; photons), and the angle the jets make with our line of sight.

Earlier versions of a unified scheme^{6–9} attached considerable importance to the last, orientation, parameter. A jet seen more or less end-on was invoked to explain several phenomena that are common, but by no means universal, in quasars, including rapid flux variability



Orientation of the approaching jet of active galactic nuclei relative to the observer's line of sight. The dividing line between quasi-stellars and radio and infrared galaxies occurs at about 44°. Thus the average radio galaxy (RG) at 69° should have an apparent, projected radio diameter about twice that of the average quasar (QSR) at 31° as seems to be the case. (Courtesy of P. D. Barthel.)

and rapid outward motion of radio-emitting blobs along the jet direction (called superluminal motion because a naive analysis of it that neglects the effects of orientation and special relativity leads to apparent motion at velocity = 2–10 c). Such a picture successfully organizes a variety of observed features of quasars and their ilk.

The classification of some quasars as roughly end-on jets and some as roughly side-on, runs into a couple of difficulties, however. First, presence of superluminal motion should be correlated with small angular size of extended radio emission as projected on the sky. It is not. The paper by Barthel *et al.*² in fact reports apparent expansion at 2.5 c for the core of the quasar 4C 34.47, which has the largest projected size of any known (560 kiloparsecs assuming a Hubble parameter $H = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$). Second, the radio flux we see should be enhanced by beaming and the X-ray flux should not; yet the two

are rather well correlated. It is also sometimes hard to understand the non-detection of the expected, diametrically opposed, jet aimed away from us.

In Barthel's restructured scheme (see figure), all quasars are beamed within 44° at us (average, 31°), and the otherwise-directed objects are seen as powerful radio galaxies, whose projected sizes are indeed larger by about a factor of two. Radio-quiet quasistellar objects may be the corresponding end-on views of galaxies that we otherwise see as strong infrared sources (because dust absorbs and re-radiates much of the emission in the equatorial direction). The presence or absence of radio emission is then an intrinsic property of a particular hole–accretion–collimation combination, and X-ray emission must be somewhat enhanced by beaming effects. The Seyfert galaxies provide some supporting evidence. The two subtypes, Seyfert 1 and Seyfert 2 galaxies, have been attributed to pole-on and equator-on directions of view on the basis of other evidence¹⁰. And, indeed, Seyfert 1 galaxies are more often the strong X-ray emitters; and the most pronounced superluminal galactic source, 3C 120, resides in a Seyfert 1. Barthel suggests that all host galaxies will be found to be experiencing some degree of interaction with neighbours. The relationship with yet another kind of activity, also widely blamed on interactions and called star bursts (meaning bursts of rapid star formation), remains to be worked out.

Barthel provides eight additional predictions of his scenario, pertaining to polarization, optical emission line, counter-jet and infrared properties of the several classes of objects. He also proposes two kinds

of observations that could rule it out — evidence for systematic differences in the morphology or in the environments between the host galaxies of quasars and radio galaxies, or between QSOs and infrared galaxies. Curiously, the definitive test of going around to the other side of a few and looking is not mentioned. □

1. Barthel, P.D. *Astrophys. J.* **336**, 606–611 (1989).
2. Barthel, P.D., Hooimeyer, J.R., Schilizzi, R.T., Miley, G.K. & Preuss, E. *Astrophys. J.* **336**, 601–605 (1989).
3. Rees, M.J. *IAU Symp.* **119**, 1 (1986).
4. Blandford, R.D. *IAU Symp.* **119**, 359 (1986).
5. Rees, M.J. *Nature* **211**, 468–470 (1966).
6. Orr, M.J.W. & Browne, I.W.A. *Mon. Not. R. astr. Soc.* **200**, 1067 (1982).
7. Scheuer, P.A.G. & Redhead, A. *Nature* **277**, 182–185 (1979).
8. Blandford, R.D. & Königl, A. *Astrophys. J.* **232**, 34 (1979).
9. Kapahi, V.K. & Saikia, D.K. *J. Astrophys. Astron.* **3**, 465 (1982).
10. Antonucci, R.R.J. & Miller, J.S. *Astrophys. J.* **297**, 621–632 (1985).

Virginia Trimble is a professor of physics at the University of California, Irvine, California 92717 and visiting professor of astronomy at the University of Maryland, College Park, Maryland 20742, USA.

Dating techniques

History revealed from bones

R.E.M. Hedges

As Mark Antony pointed out¹, most of what we should like to know about our forebears lies interred with their bones. But it is no easy task to reconstruct the cultural life of our ancestors from the fragmented evidence that their bones (and other remains) provide. Bones are very complex structures and contain potentially a wealth of information about, for example, an individual's genetic constitution, medical history, environmental influences and cultural practices. They are also far from chemical equilibrium, and information is inevitably lost with the passage of time. The effect of the burial environment, in terms of diagenetic changes both to the organic and inorganic components of bone, is even more marked, the remaining information being overwritten and further scrambled. Nevertheless, progress continues in sifting out the effects of environmental contamination and in deciphering what remains. Some new papers in *Geochimica Cosmochimica Acta*²⁻⁴ exemplify this, and reinforce research reported at a recent workshop⁵.

Analytical work on ancient bone ranges from isotope studies to immunochemical detection of antigens, taking in trace-element measurements, amino-acid racemization, protein identification and the search for DNA. So far, isotope methods promise the most straightforward results. These include dating (¹⁴C and, much more arguably, ²³⁰Th/²³⁴U), and dietary information (¹³C and ¹⁵N). But even here the state of preservation of the bone is paramount.

Radiocarbon dating of bone, especially in the critical period of the transition of man from Neanderthal to modern, thought to have ended about 30,000 years ago, can produce dates far too young because of a small admixture from younger environmental contamination. How this admixture might be separated depends on our understanding of the way in which unknown contaminant molecules can crosslink to degraded bone collagen. The same problems apply to stable isotope results; as carbon and nitrogen atoms pass along the varied food chains of man's diet, the different small degrees of isotope fractionation can be recognized. Distinctions between marine and terrestrial animals, leguminous and non-leguminous plants, as well as plants growing by different photosynthetic pathways, are easily made, and more subtle influences of environment (such as climatic stress or the effect of famine) are now being explored. But the measurements are subject to the relatively inscrutable effects of bone diagenesis.

In something of a technical *tour de force*, Stafford *et al.*⁴ have now isolated individual amino acids from bones in varying states of decay and have measured their radiocarbon age by accelerator mass spectrometry. As expected, the results are complicated, but the admixture of environmental amino acids can now be clearly followed. In a mammoth bone known to be about 11,000 years old, the dates obtained for some of the amino acids were only 3,000–4,000 years old.

The new work of De Niro and Weiner^{2,3} aims to develop strategies to recover collagen fragments free from contamination. One approach is to exploit subtleties in the biomineralization of the hydroxyapatite/collagen matrix of the bone, in which a small fraction is in the form of multicrystalline arrays. This fraction is preserved after the treatment of the matrix with sodium hypochlorite, and in fossil bone still retains isotope signatures much closer to those expected for the original bone. Non-collagenous proteins and lipids are also apparently better preserved in these 'aggregates'.

A second approach that is being explored is to take advantage of the characteristic sequential structure in collagen: glycine–Y–X, where Y is often proline or hydroxyproline and X is any residue. This structure is specifically cleaved by the enzyme collagenase into

fragments of relative molecular mass typically 500–700, which can then be extracted and purified. Although yet to be proved, the benefits for the radiocarbon dating of bone, particularly from hotter regions than northern Europe, are potentially very significant. Much Middle Eastern chronology, for example, rests on dating of unsatisfactory charcoal samples; the collagen levels surviving in bone are usually too low to provide reliable dates by existing techniques. Reliable dates obtained by these new techniques could transform our understanding of some of the archaeology of the region.

Other dating methods are less likely to evade the effects of environmental influence⁵. Uranium-series dating, for example, depends on modelling the uptake of uranium ions by buried bone from the ground water. Even the basic question of why some bones seem better preserved than others found in the same stratum on a site is not yet easy to answer. Perhaps the first steps are being taken in the continuing study of the disintegration of macromolecules from naturally weathering bones in the Amboseli National Park of East Africa. The first few years of a bone's death may be crucial for preserving its information for posterity. □

1. Shakespeare, W. *Julius Caesar* Act 3, Scene 2.2. De Niro, M. & Weiner, S. *Geochim. cosmochim. Acta* **52**, 2415–2424 (1988).3. DeNiro, M. & Weiner, S. *Geochim. cosmochim. Acta* **52**, 2425–2432 (1988).4. Stafford, T.W., Brendel, K. & Duhamel, R.C. *Geochim. cosmochim. Acta* **52**, 2257–2268 (1988).5. Hedges, R.E.M. *et al.* (eds) *Appl. Geochem.* (in the press).

R.E.M. Hedges is Director of the Radiocarbon Accelerator Unit, Oxford University, 6 Keble Road, Oxford OX1 3QJ, UK.

Circadian rhythms

Mutant hamster in a hurry

N. Mrosovsky

THE golden hamster (*Mesocricetus auratus*) is to chronobiologists what the fruitfly is to geneticists. Hamsters occupy this niche because their circadian rhythms of wheel-running activity are easy to measure, easy to quantify and highly predictable. At a recent meeting*, Martin Ralph (previously at the University of Oregon, now at the University of Virginia) reported his new work¹ on how the circadian-activity rhythms of hamsters can be experimentally manipulated by transplanting a part of the brain.

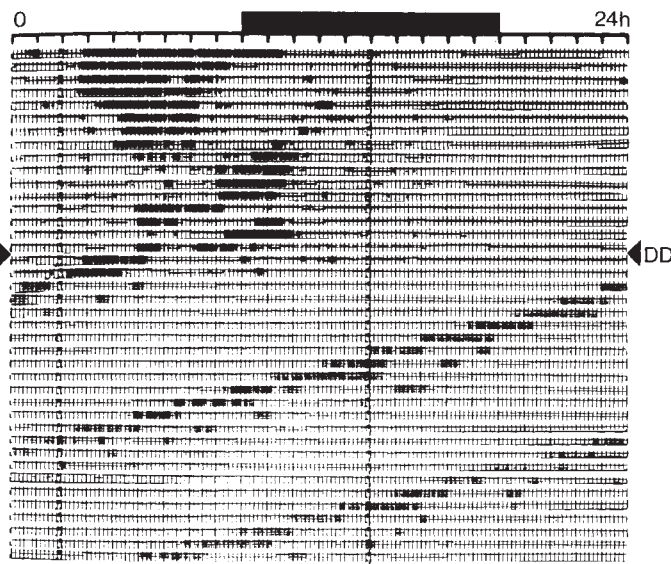
When the light–dark (LD) cycle is 14–10 hours, wheel-running in this nocturnal species is neatly confined to the hours of darkness. An occasional individual displays a sloppy or unusual rhythm, and such animals are usually discarded from

experiments. But when Ralph noticed that one hamster started to run in its wheel several hours before the onset of darkness in a 14–10-hour LD cycle (see figure), he placed the animal in constant darkness, reasoning that the early activity onset in the presence of a LD cycle might reflect a circadian pacemaker with an unusually fast frequency. In this condition the hamster's activity free-ran with a periodicity of only 22 hours. The shortest free-running rhythm previously recorded² by Ralph and Menaker, out of more than 1,000 hamsters tested, was 23.5 hours.

Tests on these offspring revealed that there is a trait, an abnormally short free-running circadian period (*tau*), at a single autosomal location³. The *tau* mutant can thus be used to tackle unresolved questions about biological rhythms. For example, it is thought that the supra-

* Society of Neurosciences, Toronto, 13–18 November 1988.

Raw data showing the discovery of the mutant *tau* hamster. Each line represents 24 hours. Successive days are mounted beneath each other. Dark pen marks indicate periods of wheel-running activity; dark bar at top, time of darkness on a light-dark 14-10-hour schedule; arrows, change of lighting schedule from light-dark 14-10-hour to continuous darkness (DD). (Courtesy of M. Ralph.)



chiasmatic nuclei (SCN) in the brain are the site of circadian pacemakers in mammals, but this assertion has not been proved. True, lesions in this area abolish circadian rhythms of activity and other variables³, but this does not prove that a pacemaker resides there. It could simply mean that the integrity of this area is necessary for the expression of rhythmicity. Transplanting fetal SCN into animals whose own SCN have been ablated can result in restoration of rhythms. This again does not prove that the SCN contain a pacemaker. The circadian machinery might be elsewhere, with the SCN simply providing some input necessary for its functioning. What is required is a demonstration that transplants produce rhythms that have characteristics which are specific to the clock of the donor.

In sparrows, removal of the pineal results in a loss of circadian rhythmicity. Menaker and his co-workers have previously shown that transplants of the pineal gland restore rhythmicity to pinealectomized sparrows, but also that the phase of the donor's rhythm determines the phase of the recipient's rhythm⁴. Thus, the time of day at which the recipient sparrows become active depends on whether the donor birds have been kept on a normal or a reversed LD cycle. This still is not quite definitive because the phase of a rhythm is a temporary characteristic that can be quickly shifted in various ways. It would be more convincing to transplant a persisting stable characteristic of a clock such as the speed with which it runs. In sparrows, however, there is too little inter-individual variation and too much intra-individual variation to make transplanting periodicity convincing.

There have been some attempts to transplant periodicity between strains of rats with different clock speeds, but activity

rhythms in rats are generally not as clear as those in hamsters. A hamster mutant with a period outside the range of those for the wild-type is the ideal subject for this kind of work.

Ralph's new data show that the period of the circadian-activity rhythms of wild-type hamsters can be shortened first by ablating their own SCN and then by inserting fetal mutant SCN. Conversely, the period of mutant hamsters can be lengthened by replacing their own SCN with fetal wild-type SCN. This result is the confirmation of the pacemaker role played by the SCN. It is also a striking example of the transplantation of a behaviour.

The mutant (*tau*) hamster is likely also to provide insight into other questions about biological rhythms. Attempts to discover the gene product of this single locus and to unravel its molecular biology have already begun. It will be interesting to make comparisons with the fruitfly period (*per*) gene, which has already been cloned and sequenced (see the recent discussion in News and Views⁵). Understanding fundamentals about rhythmicity will eventually help in ameliorating rhythm disturbances such as those occurring in jet lag, shift work and some types of affective disorder. A nice touch is that Ralph and Menaker have rejected the option of trying to patent the mutant hamster. Altogether, this is textbook science at its best. □

1. Ralph, M.R. *et al.* *Soc. neurosci. Abstr.* **14**, 462 (1988).
2. Ralph, M.R. & Menaker, M. *Science* **241**, 1225-1227 (1988).
3. Stephan, F.K. & Zucker, I. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1583-1586 (1972).
4. Zimmerman, N.M. & Menaker, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 999-1003 (1979).
5. Heisenberg, M. *Nature* **333**, 19-20 (1988).

N. Mrosovsky is in the Departments of Zoology, Psychology and Physiology at the University of Toronto, Ontario, Canada M5S 1A1.

Daedalus

A reducing diet

COOKING is not a consciously chemical discipline. Only by trial and error have chefs learnt to exploit the hydrolysis and denaturing of proteins, the partial oxidation of fats, and the Maillard reaction between proteins and sugars, to produce their wide palette of sumptuous and tempting flavours. It is largely oxygen that develops the subtleties of taste, by forming the delicate traces of higher aldehydes, ketones and esters that impart flavour; but over-oxidation gives sour, acrid fatty acids. So cooking limits oxidation: typically by careful temperature control, or (as in frying and boiling) by submersion.

Modern industrial food processing uses anti-oxidant additives; but Daedalus has a more wholesome approach. The traditional chef has to join his battle with oxygen over an open stove, and faces rancid, burnt or over-cooked disaster if he loses it. But modern electric and microwave ovens are quite independent of air. So DREADCO's chefs are cooking all sorts of dishes not in air, but in a non-oxidizing atmosphere of nitrogen or argon.

This new sort of cooking has many benefits. Very high temperatures can be explored under full control: the food cannot burst into flame. Controlled amounts of oxygen can be introduced at crucial stages, to promote the development of new and subtle flavours without going too far: just as wine develops its subtle bouquet from the slight leakage of oxygen into the bottle through the cork, but is spoiled if too much gets in. Gourmets should revel in the new and unexpected flavours that will thus be teased out of quite ordinary foodstuffs. Dull and obvious products like lentils and rhubarb and corned beef may come alive with subtle delights; hidden virtues may be revealed in marrow, tapioca, and the British sausage; the range of delightful complementary pairs, like eggs and bacon, and fish and chips, may be wonderfully extended. Culinary anaesthetics like salty gravy and sugary custard may no longer be needed. Even hopeless raw materials like lettuce and battery chicken may at last be made to taste of something.

More boldly still, the DREADCO chefs are filling their new ovens with frankly reducing atmospheres: hydrogen, carbon monoxide, even ammonia. The entirely novel flavours thus developed may be so strange and amazing as to be positively overwhelming. The hydrogen-reduction oven may be best employed as a 'decoker' to rescue overdone cooking. By converting the sour fatty acids of over-oxidation back to flavoursome aldehydes again, an over-cooked joint, deadburnt cake or pyrolysed pie might be reclaimed for the gourmet's table.

David Jones

Safety of Cameroonian lakes

SIR—In dealing with lethal releases of CO₂ from lakes, the people of Cameroon require firm scientific evidence on which to base decisions for evacuation and disaster control. We are concerned, therefore, by the misleading and speculative letter by Pourchet *et al.*¹ stating “it is likely that many of the crater lakes (in Cameroon) have emitted gas bursts, or will do so”. Our recently completed studies suggest that gas bursts from other lakes in Cameroon are, in fact, unlikely.

The accumulation and massive release of CO₂ gas from crater lakes is unique, and has been documented only for lakes Nyos and Monoun in Cameroon^{2,3}. The circumstances required for this phenomenon to occur are very rare. There must be an adequate gas supply, such as the CO₂-charged soda springs occurring in Cameroon, and this supply must be trapped by surface deposits or by a deep, strongly stratified lake. There is also good evidence that a build-up of CO₂ and solutes in bottom waters occurs long before any gas burst^{2,4}. Our recent limnological survey indicates no such premonitory build-up in any Cameroonian crater lake apart from Nyos and Monoun^{4,5}.

The argument of Pourchet *et al.* seems to rest on two points. The first is that porewaters and bottom waters of Lake Bambuluwe are “close to gas saturation at the sampling depth (58 m)”. But observing gas bubbles in a core brought to the surface, as they did, indicates little more than oversaturation at pressures (below 5.8 atmospheres) greatly reduced from those found in the bottom sediments. Such gas pressures are common in freshwater sediments; indeed nearly all of the sediment cores we have raised from 11 different crater lakes in Cameroon degassed at lake surface pressures.

The second part of their argument links the homogeneous distributions and higher than expected concentrations of ²¹⁰Pb in sediment cores from lakes Nyos and Bambuluwe. Our sediment cores from lakes Nyos and Monoun were mixed rapidly because of violent degassing at the surface. Pourchet *et al.* do not detail their methods, but only *in situ* sampling or freeze-coring would maintain stratigraphic integrity of the sediment in a gas-charged lake. In Lake Bambuluwe, the uniformly high ²¹⁰Pb concentrations they report for the upper 15 cm have other explanations. For example, a 1m core taken by us in 1985 revealed an upper 15–20-cm layer of clastic material overlying darker micaceous mud. This surface layer may correspond to the recently built access road around the lake. Such watershed disturbances are known to increase the flux of radiogenic lead to lakes⁶. We believe that the reported similarities between sediment ²¹⁰Pb in Lake Nyos and Lake Bambuluwe may be spurious, and thus inappropriate for use as predictors of future gas bursts.

GEORGE W. KLING

Marine Biological Laboratory,
Woods Hole,
Massachusetts 02543, USA

MICHELE L. TUTTLE

US Geological Survey,
Denver, Colorado 80225, USA

WILLIAM C. EVANS

US Geological Survey,
Menlo Park, California 94025, USA

1. Pourchet, M., Pinglot, J.F., Maley, J. & Melieres, M.A. *Nature* **334**, 655 (1988).
2. Kling, G.W. *et al.* *Science* **236**, 169–175 (1987).
3. Sigurdsson, H. *et al.* *J. Volcan. geotherm. Res.* **31**, 1–16 (1987).
4. Kling, G.W., Tuttle, M.L. & Evans, W.C. *J. Volcan. geotherm. Res.* (submitted).
5. Kling, G.W. *Limnol. Oceanogr.* **33**, 27–40 (1988).
6. Robbins, J.A. *Chem. Geol.* **44**, 1–348 (1984).

Dominance of South American marsupials

SIR—The discovery of mammal remains from the early Cretaceous of Cameroon by Jacobs *et al.*¹ is of unquestioned importance in interpreting mammalian evolution and biogeography, particularly regarding the key transitions to metatherian and therian ‘grades’. The statement by Jacobs *et al.* that metatherian marsupials represent “the dominant living mammals of South America and Australia”, is surely coloured by a geological perspective, however.

Although the ecological and evolutionary dominance of Australian marsupials during the Tertiary and Quaternary seems well founded, this is not true of today’s American marsupials, despite their remarkable diversity and radiations in the Tertiary^{2,3}. Modern New World marsupials comprise only three families, 15 to 19

genera, and 85 species, most endemic to tropical South America⁴. At one comparatively well-studied site in south-eastern Peru (Cochia Cashu in Manu National Park), nine species of marsupials comprise 9 per cent of the known mammal fauna — there, four mammalian orders (Rodentia, Chiroptera, Carnivora and Primates) exceed the familial, generic and specific diversities of marsupials, and Manu records of bats and rodents are incomplete⁵. Elsewhere, 34 genera and 49 species of bats have been recorded at a single locality in Rondônia, western Brazil (specimens in the Museu de Zoologia, Univ. São Paulo), and that country supports at least 7 families, 53 genera and 183 nominal species of rodents⁶.

Although 12 mammalian orders occur in South America, 43 per cent of its

mammal species are rodents, including at least 250 species of Sigmodontinae (Muridae)⁷. In terms of diversity and abundance, rodents and bats greatly outstrip marsupials as the most dominant orders of modern South American mammals. Indeed, the contrast in diversity and dominance of marsupial lineages before and after the “Great American Interchange” has been an important focus for studies of faunal dynamics and turnover.

B. D. PATTERSON

Division of Mammals,
Field Museum of Natural History,
Chicago, Illinois 60605, USA

1. Jacobs, L.L. *et al.* *Nature* **336**, 158–160 (1988).
2. Simpson, G.G. *Splendid Isolation* (Yale University Press, New Haven, 1980).
3. Marshall, L.G. *Fieldiana, Geol.* **8**, 1–65 (1981).
4. Corbet, G.B. & Hill, J.E. *A World List of Mammalian Species* 2nd edn (British Museum, Natural History, London, 1986).
5. Terborgh, J.W., Fitzpatrick, J.F. & Emmons, L.H. *Fieldiana, Zool.* **21**, 1–29 (1984).
6. Alho, C.J.R. 143–166 in *Mammalian Biology in South America* (eds Mares, M.A. & Genoways, H.H. *Pymatuning Lab. Ecol. Spec. Publ. Ser.* **6**, 1–539 (1982)).
7. Reig, O.A. *Fieldiana Zool.* **39**, 1–506 (1987).

Black holes dispute

SIR—Gunzig, Géhéniau and Prigogine¹ attempted a numerical evaluation of the inflationary model of black hole production postulated² some time ago. Unfortunately, their paper contains several errors. We emphasize that the principal hypothesis² is not destroyed as several modifications to the model of Gunzig *et al.* are not only possible, but likely.

First, the matching conditions of equations (14) and (15) of Gunzig *et al.* are stated to imply $\alpha \approx \sqrt{\beta}$. In fact these conditions are not sufficient and the value of α is very sensitive to the input. Thus, for example, as β/C^2 varies from $(1 - \epsilon)$ to $(1 + \epsilon)$, where $\epsilon \equiv 1 - \Delta_1 \approx 10^{-27}$, α varies from 0 to $\sqrt{\beta}$. Input must be fine-tuned to 27 decimal places to obtain the result of Gunzig *et al.*. It is equivalent to tuning a discontinuity in temperature between de Sitter and Robertson–Walker stages to the same precision.

Second, Gunzig *et al.* compute t_c using three helicity states. The present count, a lower bound, is 119. Using the model of Casher and Englert³ to determine the de Sitter temperature (as we consider the determination of Gunzig *et al.* unreliable) the matching radius is reduced by a factor of 10^{-23} . But the hypothesis is not destroyed because the true turnover time is expected to be much greater than t_c .

Third, Gunzig *et al.* determine the mass density of the massy component of matter from the matching condition. To obtain the number density and thence the entropy per proton they divide this by the proton mass. But the matching temperature is $\sim 10^{17} m_{\text{proton}}$. So the protons (which are certainly in equilibrium) will be in the β/R^2 term of the energy source and not

in the α/R^3 term as Gunzig *et al.* suppose. Their calculation could be used to estimate the lepto-quark density of grand unified theories, but certainly not the density of any of the present constituents of matter.

We stress that the field-theoretical foundations of the production mechanism as well as the instability of Minkowski space are flimsy, at best — though not necessarily wrong — owing to the use of an unconventional and arbitrary subtraction procedure². The model of inflation through black hole production² must be regarded as a hypothesis without solid theoretical support. If such exists, it must await the yet-to-come quantum theory of gravity. To our mind, any numerical evaluation, unless it definitely destroys the model, is premature. (A more detailed account of our calculations is available in ref. 3 or from us).

R. BROUT
PH. SPINDEL

*Service de Physique Theorique,
Université Libre de Bruxelles,
Plaine Campus (C.P. 225),
B-1050 Bruxelles, Belgium*

GUNZIG REPLIES—Brout and Spindel claim that our paper¹ contains several "errors". We firmly disagree with this conclusion. I will consider their criticisms.

The first point concerns the matching parameters α and β . We agree entirely with their remark concerning the sensitivity of α to the input. But they miss the key point, which is that there is a subtle compensation of infinitesimals which leads to our result $\alpha/\sqrt{\beta} = 0(1)$: it appears indeed that the solution of the matching equations (14) and (15) is given by $\alpha/\sqrt{\beta} = 1 - \varepsilon/\varepsilon_1$, where $\varepsilon = (\beta/C^2) - 1$ and $\varepsilon_1 = 1 - \Delta_1 \approx 10^{-27}$. It results straightforwardly from the definitions of the parameters β and C , that $\beta/C^2 = 1/\Delta_1 (T/T_{BH})^4$. The value of T/T_{BH} is easily evaluated in the euclidean parameterization of the joining conditions; the latter is appropriate as the joining point belongs to the de Sitter asymptotic region. It appears that $(T/T_{BH})^4 = 1 - \varepsilon_2$ with $\varepsilon_2/\varepsilon_1 = 0(1)$. Hence,

$$\alpha/\sqrt{\beta} = [1 - (\varepsilon_1 - \varepsilon_2)/\varepsilon_1] = \varepsilon_2/\varepsilon_1 = 0(1)$$

as we announced¹. Moreover, when the cosmogenesis model is wholly reformulated in the framework of euclidean Robertson-Walker parameterization, as we did recently⁴, the matching conditions give rise to much simpler algebraic relations than those associated with the open Robertson-Walker parameterization. These relations lead straightforwardly to conclusions identical to those resulting from the $\approx \sqrt{\beta}$ open condition.

The second point concerns the number of helicity states involved in our considerations. We mentioned this problem at the time¹. Since then, we have shown⁴

that the inclusion of any higher number of helicity states leaves all the results unaltered, provided that the black hole mass M is increased (by a factor of about 5 for $N = 100$, for example).

The third criticism concerns the use of the proton mass in the evaluation of the specific entropy. First of all, as mentioned in our paper¹, the fundamental quantity to be considered is the adiabatic invariant $(1/T)(\rho/\rho_m)$ which does not involve any elementary particle mass at all. Our calculations produce the correct experimental value for this quantity as well. Nevertheless, the quantity which is traditionally discussed in the literature is the specific entropy per proton. To obtain the latter, it is fully legitimate in our context to use the proton mass. Indeed our joining procedure connects directly the de Sitter stage to the present cosmological stage. We mentioned in our conclusions that in a more realistic model, our parameters α and β would be time-dependent.

In addition to these three specific criticisms, Brout and Spindel are pessimistic about the field-theoretical foundations of the production mechanism as well as the

instability of Minkowski space. These remarks are directed towards the series of papers on the subject, initiated by Brout, Englert and myself⁵. I share this concern. It is for this very reason that we have now developed⁴ a phenomenological approach which eliminates the field theoretical subtleties. In this context, simple thermodynamical arguments render the explicit use of field-theoretical considerations no longer necessary.

In conclusion, Brout and Spindel have missed the main points of our paper. These involve the reinterpretation of Einstein's equations leading to a burst of entropy associated with matter creation in the early stages of the Universe.

E. GUNZIG

*Service de Chimie Physique II,
Université Libre de Bruxelles,
Campus Plaine (C.P. 231),
B-1050 Bruxelles, Belgium*

1. E. Gunzig, *et al. Nature* **330**, 621–624 (1987).
2. F. Englert in *Les Houches Session XXXII* 516 (North-Holland, Amsterdam, 1980).
3. Brussels Report ULB-TH 88/05.
4. Prigogine, I. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7428–7432 (1988).
5. Brout, R. *et al. E. Ann. Phys.* **113**, 78–106 (1978).

Selective action of artificial membranes

SIR—Rubinstein *et al.*¹ claim to have prepared the first example of a stable, ion-selective artificial membrane of monolayer thickness, which successfully mimics basic structural and functional properties of natural bilayer membranes. Their membrane was prepared on a gold electrode by coadsorption of two components from bicyclohexyl/chloroform solution. After transfer to an aqueous solution, the first of these (OM) apparently acted as a blocking agent, and the second (TBEA) as an 'active' ligand forming 1:1 complexes with divalent metal ions, such as Cu^{2+} , but not with trivalent ions, such as Fe^{3+} .

To establish that their 'artificial membrane' was working as they claimed, the authors needed to prove that the only observed electrochemical reaction was caused by the active ligand chemically distinguishing between Cu^{2+} and Fe^{3+} ions in solution. Although certain evidence was adduced in favour of this hypothesis, we believe that there is an alternative and much simpler interpretation of the experimental results in which the active ligand has no role.

When the 'artificial membrane' was placed in a solution of Cu^{2+} and Fe^{3+} ions, Rubinstein *et al.* observed no Cu^{2+} 'underpotential' (non-bulk) and $\text{Fe}^{2+}/\text{Fe}^{3+}$ peaks; repression of the peak for bulk deposition of copper onto the electrode by ~ 250 mV; and a loop in the cyclic voltammetry. Rubinstein *et al.* interpret these observations as follows: Cu^{2+} peaks were absent because copper atoms were deposited inside the 'artificial membrane' at some distance from, rather than directly on, the

electrode surface; the lack of $\text{Fe}^{2+}/\text{Fe}^{3+}$ peaks was due to a lack of sites where Fe^{3+} ions could bind; and the shift in the bulk copper peak and the existence of the voltammetric loop were evidence of a "preferred perpendicular orientation of the ligand". Furthermore, they invoke quantum-mechanical tunnelling barrier to explain the shift of the copper deposition peak, and suggest a type of autocatalysis to explain the voltammetric loop.

But all the above evidence is also consistent with a simple physical model in which Cu^{2+} ions are discharged directly onto the surface of the gold electrode. The electrode may be imagined as being, say, only 99.9% covered by the monolayer and not 100% as claimed. Such a coverage would be sufficient to eliminate all but 0.1% of both the copper 'underpotential' and $\text{Fe}^{2+}/\text{Fe}^{3+}$ peaks, making them undetectable on the scale used, but would leave enough area of uncovered gold to allow the nucleation and growth of bulk copper crystals. Moreover, the high coverage would also block a large fraction of the active sites for nucleation on the gold surface, so that the nucleation overpotential for copper deposition would be increased by several hundred millivolts, precisely as observed. The loops in the voltammograms are no more than the expected response of a nucleation-and-growth process to a triangular scan of applied potential, as proved elsewhere².

STEPHEN FLETCHER

*CSIRO Division of Mineral Products,
PO Box 124, Port Melbourne,
Victoria 3207, Australia*

RUBINSTEIN *ET AL.* REPLY—Fletcher challenges our main conclusion, that the selective response of the gold electrode/(TBEA+blocking component) system to Cu^{2+} ions in the presence of, for example, Fe^{3+} is due to selective complexation of Cu^{2+} ions by TBEA molecules, and suggests an alternative interpretation. But experimental evidence¹ and other data clearly support our selective complexation model and exclude Fletcher's suggestion.

Selective behaviour of the same kind as we reported is observed when TBEA is replaced with TBPA or when the blocking agent OM is replaced with naphthol polymer (NP) (Fig. 2 in ref. 1) or polymerizable octadecyl trichlorosilane (OTS)^{3,4} (a in the figure). As we have shown previously^{3,4}, diffusion-controlled faradaic currents are generally enhanced at randomly distributed pinholes. A situation in which the bulk-copper peaks are prominent whereas copper 'under potential' and $\text{Fe}^{3+}/\text{Fe}^{2+}$ peaks are totally absent would therefore require rather unusual size and spatial distributions of pinholes. It is unlikely that three different blocking elements and two different ligands lead to the same special kind of pinhole distribution. Moreover, being deposited electrochemically, NP should block all uncoated electrochemically accessible sites. An NP layer of comparable thickness on a bare gold surface indeed blocks copper deposition completely¹.

Ionic competition experiments provide strong experimental support for the selective complexation model. Zn^{2+} and Cu^{2+} ions, which resemble Cu^{2+} in coordination requirements, successfully replace Cu^{2+} at

monolayer ligand sites although being electrochemically inactive under the present conditions. The Cu^{2+} peaks are suppressed in the presence of Zn^{2+} (a–c in the figure), and at a high Zn^{2+} concentration (~8.0 mM) the Cu^{2+} peaks cannot be detected at all. The same is observed with Cu^{2+} (in both cases the peaks are fully recovered in pure Cu^{2+} solution). Conversely, the presence of non-complexing ions, such as Fe^{3+} or Ce^{3+} , has essentially no effect on copper deposition (Fig. 2 of ref. 1; d in the figure). These results are consistent with complexation selectivity, but not with reaction at uncoated areas. Moreover, the complete disappearance of Cu^{2+} peaks in these competition experiments indicates the absence of electrochemically active pinholes down to <0.005% of the total area.

Fletcher suggests that the active element (TBEA or TBPA) plays no role in the observed selectivity. Figure 1e and f shows control experiments with pure OM monolayers on gold, in which a small number of pinholes were deliberately introduced (manuscript in preparation). At ~99.9% electrode coverage⁴, $\text{Fe}^{3+}/\text{Fe}^{2+}$ peaks indeed disappear whereas Cu^{2+} peaks still persist. However, these peaks (< 4% of the intensity observed with gold/(TBEA+OM or OTS)) clearly indicate normal copper deposition on bare gold (Fig. 2a of ref. 1), in that they show prominent copper 'underpotential' features and no unusual overpotential for bulk copper nucleation (as Fletcher suggests will be the case for pinholes). The presence of Zn^{2+} in solution here has, as

expected, no effect on copper deposition. Thus, the characteristic features of Cu^{2+} reduction cannot be observed in the absence of the active component.

Fletcher's suggestion of a nucleation overpotential rather than a tunnelling barrier does not contradict our basic model. The overpotential and voltammetric loop, indicative of a barrier to electron transfer which diminishes as metal deposition starts, can be caused by either or both of these phenomena; its precise nature has no bearing on our principal conclusions. We suggested the tunnelling barrier as a possibility based on the observation of an increased barrier when using TBPA in place of TBEA, and consequently an increased Cu^{2+} -to-electrode distance. The thickness of a TBEA monolayer — 9.5 ± 1.0 Å as measured by ellipsometry — supports the assumed perpendicular orientation of the ligand molecules, and thus the assumed distance of complexed Cu^{2+} from the electrode. Furthermore, the overpotential and loop cannot be reproduced in the absence of TBEA or TBPA.

ISRAEL RUBINSTEIN

SUZI STEINBERG

YITZHAK TOR

ABRAHAM SHANZER

JACOB SAGIV

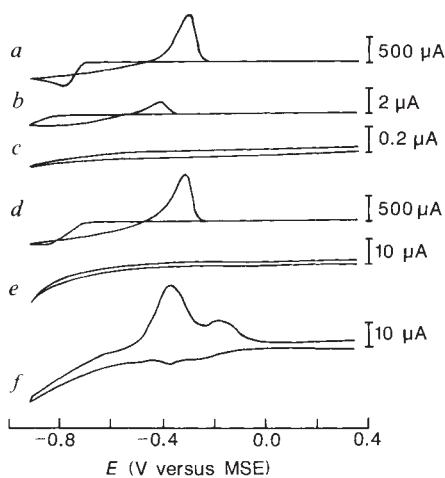
Department of Materials Research,
The Weismann Institute of Science,
Rehovot 76100, Israel

1. Rubinstein, I. *et al.* *Nature* **332**, 426–429 (1988).

2. Fletcher, S. *et al.* *J. electroanal. Chem.* **159**, 267 (1983).

3. Sabatani, E., Rubinstein, I., Maoz, R. & Sagiv, J. *J. electroanal. Chem.* **219**, 365–371 (1987).

4. Sabatani, E. & Rubinstein, I. *J. phys. Chem.* **91**, 6663 (1987).



Cyclic voltammograms in 0.1 M H_2SO_4 containing different ions (scan rate = 0.10 V s^{-1} ; electrode area = 0.63 cm^2 ; MSE is mercurous sulphate reference electrode, +0.400 V versus saturated calomel electrode). a, Au/(TBEA+OTS) in 1.0 mM Cu^{2+} ; b, Au/(TBEA+OTS) in 1.0 mM Cu^{2+} + 6.0 mM Zn^{2+} ; c, Au/(TBEA+OTS) in 1.0 mM Cu^{2+} + 8.0 mM Zn^{2+} ; d, Au/(TBEA+OTS) in 1.0 mM Cu^{2+} + 6.0 mM Ce^{3+} ; e, Au/OM with controlled pinholes in 3.0 mM Fe^{3+} ; f, same as e in 1.0 mM Cu^{2+} .

Rapid reversibility of lake acidification?

SIR—Battarbee *et al.* recently claimed¹ that diatom cores showed "a trend towards progressively decreasing acidity" in two Scottish lochs. They put these data forward as evidence for rapid reversibility of acidification in Galloway lochs in response to a decrease in acid deposition, in contrast to model predictions which show a much slower response. Unfortunately, the data do not justify this optimistic conclusion, although they do raise some interesting questions.

Battarbee *et al.*¹ do not present the customary diagram of pH change with time on the grounds that the changes observed are within the errors of the model for pH reconstruction. The diagram can be constructed (Fig. 1) from measurements on their Fig. 2 (which incidentally contains a mistake: the percentage frequency figures would be multiplied by 2) using the index B method described elsewhere². There is clearly no simple relationship between the inferred pH of Loch Enoch and sulphur deposition at Askalemuir nearby³, or between inferred pH and SO_2 emissions in the United Kingdom. A reasonable inference would seem to be that the pH of Loch

Enoch is not sensitive to short-term (10-yr) changes in sulphur emissions of up to 40%, within the limitations of the diatom core method ($\pm 0.3 \text{ pH units}$).

To reduce the uncertainty of the diatom methods, Battarbee *et al.* used correspondence analysis, an ordination technique which aims to simplify the structure of multivariate data by projecting it onto several principal axes⁴. The difficulty with such techniques is that the axes have no direct physical integration: this has to be inferred from other evidence. One criterion for a successful ordination is that replicate samples show the same patterns⁴. The two-dimensional projections of Battarbee *et al.*¹ show similar patterns only in that the general direction for replicate cores in each loch is the same. However, in one loch the time track goes east–west. Recent changes in direction are meant to indicate recovery, but in Loch Enoch the time track which had been proceeding south–east turns west and then north, whereas in Round Loch of Glenhead the 1984 core track describes a little circle and turns north–west, and the 1986 core which was going west–north–west turns south–east.

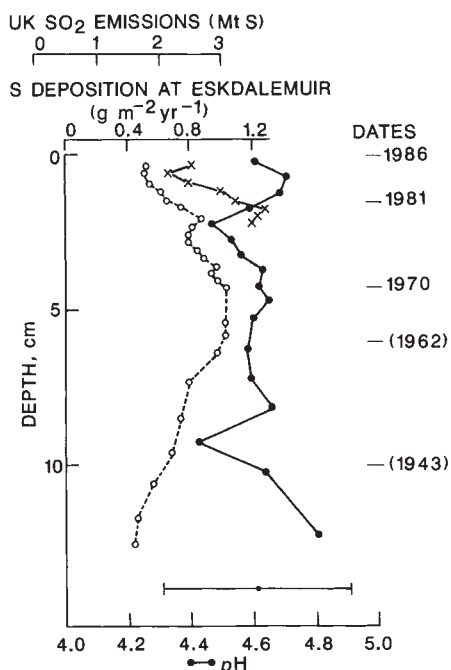


Fig. 1 Inferred pH (index B)² in Loch Enoch from Fig. 2 of ref. 1 (closed circles). Dates are taken from ref. 1 and, by inference, from the 1982 core (in parentheses). The estimated standard error of the method² is shown below. Sulphur deposition at Eskdalemuir³ 80 km distant (crosses) and UK SO₂ emissions³ (open circles) are also shown.

It is difficult to see how such diverse behaviour implies acidification followed by recovery, particularly as the three tracks in Fig. 4 of ref. 1 occupy different positions in two-dimensional space.

The chemical evidence presented by Batterbee *et al.*¹ appears more compelling, particularly as the 'non-marine' sulphate has decreased roughly in parallel with deposition, implying that these catchments do not store or release sulphur. However, only two or three samples per year were taken and there have clearly been influences on water chemistry other than sulphur deposition. The 45–60 per cent increase in Cl⁻ between the 1978 and 1984 samples shows there were one or more sea-salt episodes influencing the chemistry of the later period. The non-marine calcium has doubled in one loch, a result completely at variance with correct understanding of acidification and deacidification processes⁵, which would predict a decrease in Ca²⁺ with reducing SO₄²⁻. Has there been some increase in calcium input to these catchments or some catchment disturbance (which would also account for the increase in sedimentation rate¹)? I conclude from the paper¹ that diatom-core data have no rapid recovery resulting from reduction in SO₄²⁻ deposition, but that direct chemical data show intriguing patterns for further investigation.

RICHARD SKEFFINGTON
Central Electricity Research Laboratory,
Kelvin Avenue, Leatherhead,
Surrey KT22 7SE, UK

BATTARBEE *ET AL.* REPLY—Skeffington's attempt to dismiss our evidence for the reversibility of lake acidification seems to be based on a misreading or misunderstanding of our paper¹. First, we do not argue for a rapid recovery, rather for "little delay ... in the response ... to a decrease in acid deposition". The extent of the recovery is quite minor.

Second, we argue that the sensitivity of diatom-pH reconstruction models is too low to be of value in this context where the change is about 0.1–0.2 pH units, yet Skeffington presents such a reconstruction.

Third, we argue for floristic reversal as a more sensitive indicator and the ordinations presented clearly show that present

the pH of surface waters. Therefore, contrary to Skeffington's implied argument that the higher sea-salt concentrations in the 1984–85 data set would reduce the pH difference between the sampling periods, the differences would have been greater if the 1978 values had been maintained.

Sixth, it is not relevant to use the usual sea-salt ratios to calculate non-marine calcium because cation exchange processes can also produce elevated calcium levels as well as lower pH values in surface waters with sodium being retained on soil-exchange sites⁹.

Contrary to Skeffington's view, both the floristic and chemical data presented in our paper¹ can only be interpreted as

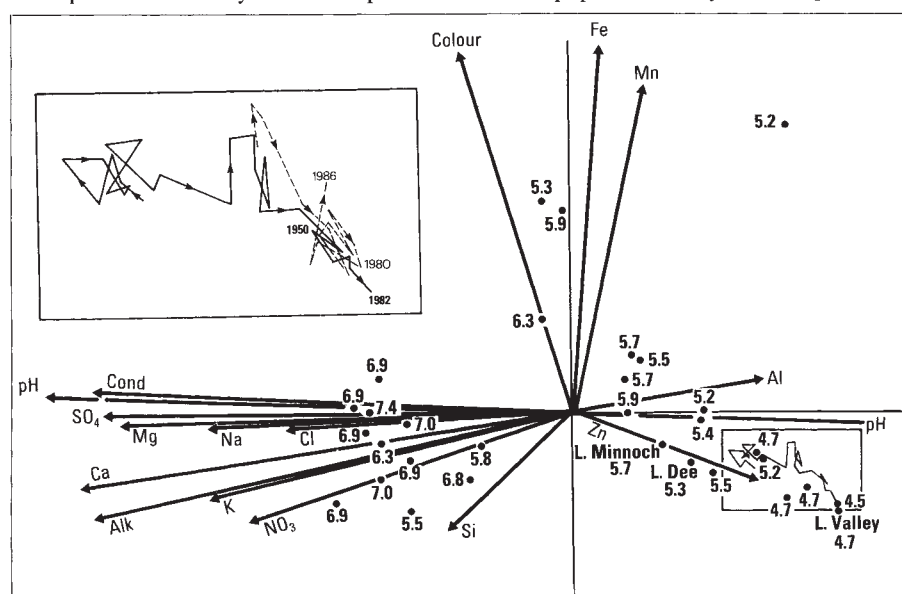


Fig. 2 Canonical correspondence analysis of Galloway lakes showing environmental axes and the pH of sites. The time-track for the 1982 Loch Enoch core is also shown. The inset shows an enlargement of this time-track along with the time-track for the 1986 core.

diatom assemblages are similar to pre-1970 floras. It is not necessary to show that the axis is a pH one as the acidification histories of both these sites are already known^{6,7}. To demonstrate this point, we use in Fig. 2 a canonical correspondence analysis⁸ of the Loch Enoch data as before, but here constrained by the modern chemistry and diatom assemblages of other Galloway lakes. The first axis is clearly identified as a pH-related axis.

Fourth, Skeffington has misread the frequency of chemical sampling. Samples were taken at 2–3-month intervals, not 2–3 times per year.

Fifth, increases in sea-salt inputs reduce

evidence for reversibility. It would be useful to assess whether similar responses have occurred at other sites with differing catchment characteristics, and to consider the extent to which reversals can be sustained if sulphate deposition were not further reduced¹⁰.

RICHARD W. BATTARBEE
ROGER FLOWER
TONY STEVENSON

Palaeoecology Research Unit,
University College London,
26 Bedford Way,
London WC1H 0AP, UK

RON HARRIMAN

Freshwater Fisheries Laboratory,
Pitlochry, Perthshire PH16 5LB, UK

1. Batterbee, R.W. *et al.* *Nature* **332**, 530–532 (1988).
2. Batterbee, R.W. *Phil. Trans. R. Soc. B* **305**, 451 (1984).
3. Barrett, C.F. *et al.* *Acid Deposition in the United Kingdom 1981–1985* (Warren Spring Laboratory, Stevenage, 1987).
4. Gauch, H.G. *Multivariate Analysis in Community Ecology* (Cambridge, 1982).
5. Reuss, J.O. *et al.* *Nature* **329**, 27–32 (1987).
6. Flower, R.J. & Batterbee, R.W. *Nature* **305**, 130 (1983).
7. Flower, R.J. *et al.* *Journal of Ecology* **75**, 797–824.
8. Harriman, R. & Wells, D. *Journal of Water Pollution Control* **84**, 215 (1985).
9. ter Braak, C.J.F. *Ecology* **67**, 1167–1179 (1986).
10. Neal, C., Whitehead, P. & Jenkins, A. *Nature* **334**, 109 (1988).

Whale correction

In the letter¹ by K. Ralls and R. Brownell Jr, refs 4–6 were omitted from the reference list. They are reproduced below.

1. Ralls, K. & Brownell, R.L. Jr *Nature* **336**, 116 (1988).
4. Jones, M.L. & Swart, S.L. in *The Gray Whale Eschrichtius robustus* (Academic, Orlando, 1984).
5. Brownell, R.L., Jr & Ralls, K. *Rep. int. Whal. Commun. (Spec. Iss. 8)*, 97–112 (1986).
6. Smith, R.L. (ed.) *Sperm Competition and the Evolution of Animal Mating Systems* (Academic, Orlando, Florida, 1984).

The inquiry inquiry

Wolfgang Rüdig

Sizewell B: An Anatomy of the Inquiry. By Timothy O'Riordan, Ray Kemp and Michael Purdue. Macmillan, London: 1988. Pp.474. £45.

EVER since the Windscale inquiry of 1977, the public inquiry has been central to any discussion of the politics of nuclear power in Britain. The 100 days of Windscale made the 'civil' use of nuclear power a prominent issue in British politics, and debate over the virtues and drawbacks of the 'ritual' of nuclear inquiries produced countless articles and books. When in 1980 the government announced its intention to hold a public inquiry into the first pressurized water reactor (PWR) to be built in Britain, the stage was set for further argument about public involvement in reaching decisions about nuclear power.

The inquiry into the PWR project at Sizewell in East Suffolk eventually opened in July 1982. It broke all records in terms of its length and costs, finally being closed in March 1985 after sitting for 340 days. The inspector's report was formally submitted to the responsible minister in December 1986; his consent for the project to go ahead was given in March 1987. Work started in the same month, more than seven years after the respective nuclear-expansion programme had first been announced by the government. But by then, there was no powerful opposition movement making a determined effort to stop it. Although the political fall-out of Chernobyl had swept away most other nuclear development programmes in Western Europe, Britain was able to start a new phase of nuclear expansion without any serious difficulties over legitimization. Any analysis of this situation has to take the Sizewell B inquiry into account.

A team from the University of East Anglia, led by Professor Timothy O'Riordan, was given the task by the Economic and Social Research Council of analysing the inquiry. The results of the project have now been published. Mirroring its subject matter, *Sizewell B: An Anatomy of the Inquiry* can claim to be the longest and costliest book ever written on a British public inquiry. After five years of research, which was completed in May 1987 and has resulted in this well-produced book, we may ask "Was it worth it?". Do we now know more about the Sizewell B inquiry, the reasons for its length and costliness, its result and the political repercussions? Are we now better placed to evaluate the political conflict over nuclear energy and assess the effect of the inquiry on nuclear policy making?

The authors certainly faced a mammoth task in coming to grips with the unprecedented amount of material generated by the inquiry. They help the reader by including a number of appendices providing, *inter alia*, information on the chief parties involved and an indispensable list of acronyms (a total of 159). Armed with such auxiliary implements, one can delve into the maze of 12 chapters, covering a further 413 pages, which examine the inquiry from different angles.

Was it worth it? On the positive side, the description of the events surrounding a number of key issues is a notable

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Weight of evidence — 2½ hundredweight of it at the Sizewell inquiry from the Central Electricity Generating Board alone.

achievement. The arduous routine of attending endless inquiry sessions and meetings, and working through the heaps of papers involved, deserves credit which should go particularly to the senior research associate, Dr Ray Kemp. On the analytical front, it is the treatment of the 'internal' features of the inquiry, such as the innovative introduction of an inquiry counsel, the 'top table', and the 'side-room meetings', which is worked out especially convincingly. Another strong point of the book is the coverage of the legal background.

The aims of the authors, however, go far beyond the provision of a descriptive account of the inquiry with limited examination of its legal aspects and the roles of the individual actors. They set out to give an authoritative account and a comprehensive analysis, explaining the course and the result of the inquiry, and to provide recommendations for the future. It is

measured against these more ambitious claims that the book must be considered a failure.

No integrated theoretical framework is developed to analyse the inquiry. Some individual snippets of social science theory are alluded to, but reference to them is patchy and marginal. The questions raised by these theoretical points are pursued separately, so much of the analysis remains rather disjointed. We find a normative-legalistic evaluation of criteria of fullness, fairness and thoroughness going hand-in-hand with attempts to see the inquiry as a means to gain legitimation for a predetermined decision. Both approaches also have to compete against an implicit policy orientation, which is never developed theoretically, analysing the inquiry essentially as an administrative tool.

As a result, there are some glaring contradictions. For example, the reader is told fairly early in the book that the outcome of the inquiry was "inevitable" because the "technology, its supporting organizations and the people involved were just too powerful, too big and too committed to be deflected" (p.39). One chapter later, however, we learn that the public inquiry cannot just be regarded as a place for objectors to let off steam, but that it is an "extra-parliamentary device for political debate where conflicts of interest over principles and priorities of natural import can democratically be played out" (pp.52-53). No explanation is furnished of the "democratic" nature of such an exercise. Further, there is a range of suggestions for future public inquiries with the aim of making them, amongst other things, more "efficient". In the light of the previous thesis on the "inevitability" of the outcome, one is bound to ask who would benefit from more "efficient" ways of conducting such inquiries.

O'Riordan and his colleagues see the main result of the Sizewell B inquiry as being its lasting effect on the management of nuclear projects and the conduct of the licensing process. This aspect of the inquiry's ramifications is well worked out and convincingly argued. But, apart from rationalization of the nuclear sector's own procedures, has the inquiry managed to gain legitimacy for a nuclear project whose inevitability the authors consider to have been established before and outside the inquiry? On this point, the book is very weak. While it is said that the factors determining the outcome of the inquiry are to be found outside the inquiry itself, no effort was made to investigate them more fully. Scant attention is given to the state of the nuclear industry and the anti-nuclear movement. Although the authors claim to have conducted intensive interviews with "key participants and observers" (p. viii) there is no further reference to these interviews in the book, and no

clear-cut judgement of the inquiry's influence on the future of the nuclear industry and the anti-nuclear movement in Britain. And although there is some assessment of the possible long-term implications of changes to nuclear project management, a discussion of the future public acceptability of nuclear power is absent.

Was the fullness and thoroughness of the inquiry effective in diverting opposition groups from potentially more effective means of protest? What happened to the groups that boycotted the inquiry because of its alleged unfairness? How can we explain the combination of a perceived

unfairness, both in terms of the imbalance of resources between proponents and objectors, and the alleged predetermination of the outcome, with the apparent lack of public response to its conclusion? These questions are not seriously addressed. The book provides a thorough and authoritative account of the inquiry proceedings, but the final evaluation of the effects of the Sizewell B inquiry has yet to be made. □

Wolfgang Rüdiger is a Lecturer in the Department of Politics, University of Strathclyde, 16 Richmond Street, Glasgow G1 1XQ, UK.

Immune response

J.R. Batchelor

Milestones in Immunology: A Historical Exploration. Compiled by Debra Jan Bibel. *Sci Tech Publishers, 701 Ridge Street, Madison, Wisconsin: 1988. Pp.330. \$35. Distributed outside North America by Springer-Verlag, DM64, £21.*

COMPILING an anthology of scientific papers judged to be milestones in a subject can be a stimulating occupation on a winter's night, particularly in company with argumentative friends. Of course there will be areas of agreement as to what should be included and who made a particular observation, but strong disagreements are also likely to be voiced. Debra Bibel's *Milestones in Immunology* is therefore a provocative as well as an informative book.

To our shame, few of us read original articles published more than five or ten years before we started work on our own doctoral theses, particularly if they are not concerned with the limited fields in which we are engaged. This failing accounts for much of the dull incoherence and the narrow perspectives to be found in the introductory historical chapters of those same theses. Dr Bibel's book provides a partial solution. She has selected her 'milestone' papers, grouped them loosely into nine sections (immunotherapy, immunochemistry, cells and interactions, and so on), provided long excerpts of the papers, and accompanied them with commentaries on both the science and authors.

There is much to enjoy here, according to individual tastes. For example, I found it interesting to be reminded of the extraordinary bureaucratic resistance to the introduction of prophylactic immunization against typhoid for British soldiers at the beginning of the twentieth century, a measure advocated with much spirit by Almroth Wright. There is the concise logic of Linus Pauling's paper on the structure of antibodies in which he assumes,

with Marrack, Heidelberg and others, that because antibody can form a precipitate on combining with antigen, the antibody molecule must have two combining sites. "The rule of parsimony (the use of the minimum effort to achieve the result) suggests that there are only two such regions, that is, that the antibody molecules are at the most bivalent". The fact that there is an exception to this rule, and that the 'instructive' theory of antibody formation he proposed in the paper is wrong, does not detract from the interest. There are, too, some apt or ironical quotations introducing the chapters. For example, the line from Edmund Burke ("Our antagonist is our helper") or that from Paul Anderson ("I have yet to see any problem, however complicated, which when you looked at it in the right way, did not become still more complicated").

In spite of many attractive features of the book, there are aspects which I found disappointing. To a large extent they are inherent in treating a scientific enterprise as if it consisted of a series of pedestrian studies, too dull to be noteworthy, punctuated by the inspired contributions of a few outstanding people.

Dr Bibel will not escape the charges of oversimplification and, at times, even injustice. To select one of G.D. Snell's papers describing a strategy for identifying the alleles of minor histocompatibility system genes in a series of congenic mouse strains, and use this as a reason for placing his work in the section on technology, is admitted by the author as likely to invite criticism; but it mistakes a branch for the whole tree and those who do not already know of the studies of Snell and his colleagues would not, from this book, be able to appreciate their theoretical significance. In my view, the importance of C.C. Little's hypothesis, formulated early this century — that progressive growth of transplanted tumours depended upon graft and host sharing multiple, independently segregating genes — should have been recognized. The hypothesis provided the motive for the development of inbred strains, which were needed to test Little's theory. Snell's great achievement

was that he placed these genes into separate linkage groups, determined the extent of their polymorphism and defined their relative influence upon graft survival. This achievement led to the concept of each species having a single major histocompatibility system and multiple minor systems, as well as providing many invaluable mouse strains as tools for research.

In the description of Peter Gorer's work, Dr Bibel comments briefly:

His new finding, however, was the correlation of resistance to tumor transplantation to the absence of a particular blood group antigen. He offered tumor studies an immunological perspective, and provided the first linked marker of histocompatibility.

Few would guess from this remark how clearly Gorer's experiments had shown that the fate of a transplanted tumour depended on whether or not the recipient mice mounted an immune response against antigen II expressed by the tumour. And it seems strange that after Gorer had thus demonstrated the immunological basis of graft rejection, Dr Bibel does not cite his pithy conclusions:

Normal and neoplastic cells contain isoantigenic factors which are genetically determined. Isoantigenic factors present in the grafted tissue and absent in the host are capable of eliciting a response which results in the destruction of the graft. Under special circumstances the response may not be elicited or grafted tissues may not be destroyed thereby.

I also miss acknowledgement of the work of so many scientists who deserved to be mentioned. For example, in the commentary on Jean Dausset's discovery of human leukocyte alloantibodies, and the analysis of the HLA system carried out at the international workshops, there is no recognition of how the success of these analyses depended critically on the solution of at least three problems. First, van Rood showed how to recognize antisera containing antibodies of the same specificity, using computers to perform the large number of χ^2 tests required. Second, Walford and colleagues discovered that rabbit complement allowed Gorer's lymphocytotoxicity test to be used with human reagents, thus providing a reliable test. Third, Terasaki and co-workers invented the 'micro' version, thus permitting small amounts of defined antibody reagents to be used by many laboratories for standardization purposes. None of these was a small achievement. Each is an example of how misconceived it is to imagine that progress in understanding depends upon the efforts of a few; the reality is that it is due to the applied talents of the many. □

J.R. Batchelor is a Professor in the Department of Immunology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK.

Life and loves of a Martian pioneer

Peter Kemp

Genesis: An Epic Poem. By Frederick Turner. Saybrook, 4223 Cole Avenue, Dallas, Texas 75205: 1988. Pp.303. Distributed by W.W. Norton, pbk \$9.95.

INSPIRED by such varied sources as Homer, James Lovelock's Gaia hypothesis, the Bible and the Wisconsin Arboretum's success in turning degraded farmland into prairie, Frederick Turner's *Genesis* is a work of considerable compass. Aiming to synthesize old and new, art and science, it tells in epic verse the story of the colonizing of Mars: as the settlers contrive to transform an arid planet into garden, Turner seeks to re-fertilize an apparently defunct genre.

His task is as risky as it is ambitious and, conscious of this, he equips his epic with defence mechanisms. This 10,000-line saga is, we are informed, the work of a twenty-first-century poet living under mild-ish totalitarian conditions then prevailing on Earth. A martyr to arthritis and despondency, he is painfully aware that his verse lacks polish. Also, we are warned that in the process of projecting it into Turner's cortex — exactly how is not explained — clumsiness may have crept in. Further intimating that the poem should be handled with kid gloves is a heartfelt evocation — "Poor science-fiction. Last muse of the gods . . ." — lamenting the unjustly despised status of this type of writing. As a final deterrent against criticism, the poem features as its most repulsive character a reviewer who has "A Shiite mullah's openmindedness, / A moral backbone of boiled watercress; / All the prophetic vision of a sheep".

At risk of seeming such a mix of bigotry, salad and mutton, it has to be said that the poem is uneven. Its best stretches are its landscapes. Turner sensuously excels at evoking atmospheric terrains — from the wind-scorched Australian outback to Mars at different stages of its ecogenesis: the "golden furs and powdery crimsons" of mosses and lichens spreading over "glitterings of slime and foams", blossoms gorgeously exploding into newly oxygenated air.

But, along with this flair for dramatizing the processes of life-formation and evolution, goes an enfeebling inability to create and develop lifelike characters and situations. The poem is peopled by melodramatic puppets. Resembling the plastic personnel of some *Star Wars*, they range from evil Gaea, perfidious leader of the Ecotheist Movement, which teaches that nature must never be interfered with, to noble Chance, her estranged husband,

dedicated to bringing life to Mars. As battle-ships, a space Ark and a displaced ice moon of Saturn hurtle round the Solar System, freedom fighters, visionaries, gladiators, lurid villains and a miracle-working prophetess engage in comic-strip adventures.

Bizarrely, Turner's celebration of advanced biotechnology is studded with throwback violence: hand-to-hand fights, karate chops, disembowelling. Martial arts loom as large as Martian pioneering in his story. Incongruity keeps breaking through the attempt to combine sophisticated scientific lore with the barbarities traditional to epic literature: wars, parricide, split families, blood feuds, vengeance.

The poem's classic in *medias res* opening finds an appropriate out-of-this-world parallel to the *Iliad*: helmeted figures in metallic uniforms clash on the deadly plains of Mars. But despite the subsequent efforts to elevate proceedings by using the apparatus of the epic — heroic similes,

addresses to the muse, cataloguings of troops — bathos often intrudes. After an assassin returns from a mission to kill the Sibyl at her shrine in the Peacock mountain on Mars, it comes as an anti-climax to hear of his booking into a "clean hotel just west of Reading". Behaviour oscillates between the cosmically marvellous and the comically mundane. The switches of language also jar, with abrupt changes from laboratory terminology ("volatile calyptal ribosomes", "synthesized endorphins", "heterodyning frequencies") to musty archaisms ("And ruddy Phoebus gins to welke in west"). And it seems a shame that the twenty-first-century poet, though familiar with the intricate mechanisms of planetary engineering, has not always worked out ("tech-/Nical", "psych-/Ologically", "choc-/Olate") how to fit his words neatly into pentameter lines. □

Peter Kemp, 61 Princes Avenue, Finchley, London N3 2DA, UK, is lead fiction reviewer for the London Sunday Times.

We've got rhythm

Asa Briggs

The Metronomic Society: Natural Rhythms and Human Timetables. By Michael Young. Thames & Hudson/Harvard University Press: 1988. Pp. 301. £16.95, \$25.

MICHAEL YOUNG, whose book *The Rise of the Meritocracy* fascinated readers a generation ago, has produced another remarkable and equally stimulating book which deals with even more basic issues. It is designed to be read by a wide audience, and it abounds in memorable phrases and descriptions. "Durability is not durability unless it is pulsating." "Habits are not usually chosen with any deliberation; they just grow, wild flowers rather than cultivated ones." "A habit is a memory unconsciously edited for action." "The fertility of nurture festoons itself around nature." "Time is immanent in everything . . . You cannot catch it because you already have it." In the broadest sense this is a book for everybody, rich both in analogies and in insights. It deals in various degrees of detail with such topics as time, memory, tradition, habit and custom. It begins with an analysis, continues with a set of explanations and ends with a prescription.

The term "metronomic society" is not introduced until the end of the first chapter, where we are told that "modern society (and social science with it) . . . has a linear bias to it, and that with this linear bias many natural rhythms have been replaced by artificial ones, a rhythmic society replaced by a metronomic". The second chapter, which begins by noting how the living organism has long been one

of the favourite analogies for society, examines more fully the nature of the natural rhythms. It summarizes research, not all of it completely convincing, on an internal biological clock, while adding that it will not be conclusive until the clock or clocks shall be found.

Lord Young then turns to social behaviour and how it is timed, drawing the necessary contrasts both between natural rhythms and artificial rhythms and between animals and human beings: "bees and birds do not celebrate Sundays". Regularities, both natural and artificial, "create the social mesh", but "particularly within large and complex organisations, biological and social rhythms are in conflict with each other". The fourth chapter, "Habit: the Flywheel of Society", is the most original, although the term habit is used very generally to cover phenomena which are only differentiated at the end of the chapter (via the *Encyclopedia of Social Science*) into motor habits, cognitive habits, emotional habits and moral habits. It would have been more persuasive if the differentiation had been introduced from the outset, for the various activities that Lord Young uses as examples of habit suggest differences more than similarities. The conclusions are firm, however — "Habit is a means of making learning stick"; "Habits constitute the rules of individual behaviour and customs the rules of social behaviour".

There is a darwinian ring to the last paragraph of the fourth chapter, which sets the scene for the following discussion of "Social Evolution". This is a less original chapter. It deals with speech, law and institutions, taking the British monarchy as the example, and notes how "despite and because of the hold of habit, institutions can evolve" and how in each

case "change comes not by a wholesale overthrowing of the past but by leaning against the past to gain a kind of leverage and add something new to it". A glance at another Young, G.M. Young, might have been useful here. Even so, Lord Young, who draws most of his analogies from biology, has some useful comments to make about history. In comparison economics throughout gets short shrift, even in the brief section "The Question of Progress", and even though economists have been particularly interested, as the author observes *en passant* earlier in the book, in cycles and in 'trends' (a term he does not use).

Nonetheless, Chapter 6, "The Metronomic Pulse", begins with economics although it is approached ritualistically rather than analytically, and later turns to a point made by a number of economists that "the axiomatic and automatic consequence of a rising standard of living is that time becomes scarcer". For Lord Young the position is more serious than they suggest, for "the symptoms of the famine are all around us" and (a memorable phrase) "the more deliberation there is, the more deliberation is necessary". The thought leads, in linear fashion, to the prescriptions of the last chapter — reintroducing "more of the cyclical into the evolutionary mix" and giving "a more prominent place once again to the cycles of nature". The penultimate section is called "Back to Circadian Rhythms", the last "Escape from the Grid". The titles speak for themselves. To anyone suffering from time famine the prescription is attractive, and there may be reasons different from those advanced by Lord Young for following it. He does not quite say, however, how to do so.

Within the book there are running arguments with sociologists about the nature of sociology, at least one dig at anthropologists, and freely acknowledged processes of borrowing from biologists. Indeed, while expressing distaste for sociobiology Lord Young shows himself to be an eloquent spokesman for his own version of social darwinism. He has great social concern, however, and is untouched by determinism. One word he never uses about human beings is curiosity — perhaps because he has so much of it himself.

Lord Briggs is Provost of Worcester College, Oxford OX1 2HB, UK.

New in paperback

- *Medicine in China* by P.U. Unschuld. Publisher is University of California Press, price is \$12.95, £7. For review see *Nature* 327, 198 (1987).
- *The Great Devonian Controversy* by Martin J. S. Rudwick. Publisher is University of Chicago Press, price is \$22.95, £15.95. For review see *Nature* 316, 491 (1985).
- *Duchenne Muscular Dystrophy* (revised edition) by A. E. H. Emery. Publisher is Oxford University Press, price is £17.50, \$35. For review see *Nature* 329, 594 (1987).

Infant individuals

George Butterworth

Nature and Nurture During Infancy and Early Childhood. By Robert Plomin, John C. DeFries and David W. Fulker. Cambridge University Press: 1988. Pp 345. £30, \$44.50.

IN THIS book, Robert Plomin and his colleagues employ quantitative behavioural genetics to explore the origins and development of individuality during infancy and early childhood. They analyse "the nature of nurture" to study genotype-environment interaction and correlation, and the genetic mediation of environmental influences, rightly stressing that their methods yield totally different information to that obtained by research on group averages. The second approach consigns individual differences to error variance, whereas the first respects natural variation as being at the very foundations of individual development.

Behaviour genetics involves detailed and rather technical statistical analyses of twin and adoption studies. Twin studies, in which identical and fraternal twins are contrasted, enable estimates of the components of variability due to genetic or environmental effects. Adoption studies, on the other hand, allow the possible covariation between genes and environment to be teased apart. Genetically unrelated individuals adopted into the same family will resemble each other for environmental reasons only. The effects of the remainder of the non-shared environmental variance, such as differential maternal attention between siblings, can also be estimated.

Height and weight, obvious physical variables, are often used to 'anchor' cognitive and behavioural measures. Height, for example, shows a strong correlation from infancy to early childhood, with genetic and environmental effects differing at various times in development. Fraternal twins are actually physically more alike than identical twins at birth, but not at six months, perhaps because identical twins may be subject to greater intra-uterine competition for nourishment through the chorion. Adopted siblings also show significant correlations for height during the second year of life, although these values are not as high as for fraternal twins. Hence, for physical variables it is possible to distinguish genetic effects, effects of the shared twin environment and effects of non-shared environmental factors.

Cognitive or personality development is, of course, more variable than physical development. IQ scores are notoriously unstable from infancy to early childhood, at least in part because of the different sensori-motor and verbal items on the

tests. Genetic influences on IQ may be greatest between the ages of one and two years, while shared environmental factors, as assessed through adoptive siblings, may account for more of the variability between IQ in early childhood. The Colorado Adoption Project (CAP) analysed here also shows that shared environment may be of lesser importance for the development of temperamental characteristics than non-shared environmental effects. An odd, unexplained result of the CAP study was that the amount of television viewing at three and four years of age may have a heritable component; a significant correlation was found on this variable between adoptive siblings.

In the final chapter, the authors consider the "nature of nurture". The appropriate behavioural genetic design for such a study would involve quantifying the parenting behaviour of adult identical and fraternal twins as well as data from twin children and their parents. Genetically mediated differences in parenting, as revealed by differences between identical and fraternal twin parents, may have differential effects on their children. Genetic differences among children may also contribute, and the differential effects could be quantified because the children of identical twins are related to each other as half-siblings, whereas the children of fraternal twins are cousins. Parents and children share 50 per cent of their genes, regardless of whether their twin children are identical or fraternal, and this again yields the possibility of contrasting differential effects on development. Unfortunately, there are no data on these contrasts; in the book they seem to be mentioned as a behaviour geneticist's methodological ideal. But adoption studies do offer some evidence that maternal involvement — in encouraging development, for example — correlates more highly with IQ in non-adoptive siblings than in the adoptive child. The authors suggest that this may be explained as a genotype-environment correlation whereby the greater genetic match between the mother and her biological child facilitates development.

The book contains descriptions of many hundreds of observations on the complex inter-relationships between heredity and environment, and they illustrate well the ingenious ways in which the components of individual variation can be disentangled. The authors have made a great deal of progress beyond the stark genes versus environment dichotomy of earlier behaviour genetics. Whether the greater sophistication of partitioning will satisfy those who espouse a holistic approach to individual development remains to be seen; but at least the inter-penetration of nature with nurture has been acknowledged. □

George Butterworth is a Professor in the Department of Psychology, University of Stirling, Stirling FK9 4LA, UK.

Graduates — for better, for worse

J.M. Ashworth

Should a university be devoted solely to teaching? Data from the University of Salford suggest that such an institution would be unlikely to produce good graduates.

BRITISH universities receive grants from the state through the University Grants Committee (UGC) which they use to support both the education of undergraduates and the research of their staff. The British government has suggested that universities should account separately for what they spend on research and on teaching, which has given rise to a debate on the extent to which the quality of undergraduate teaching requires that those doing the teaching should also be carrying out research.

In a recent report¹, the Advisory Board for the Research Councils (ABRC) recommended that British universities should be classified into groups, with one group (type T) being defined as those that are unable to provide comprehensive facilities for research. The government has rejected this advice, but many fear that such a policy may be introduced by stealth as a result of the current UGC reviews of university departments of subjects such as physics and chemistry. The UGC will be replaced by the new University Funding Council (UFC) in April 1989, and it is this body that will have to decide whether or not to implement the recommendations of the reviews.

Lack of data

The debate has been marked by a lack of data. Those who believe that high-quality teaching can be carried out only in a university which is also a research centre have been placed at a disadvantage by the fact that, although it is relatively easy to adduce anecdotal evidence for the proposition that 'not all good researchers are good teachers' (or vice versa), it is not easy to demonstrate that good teaching depends on good research. It is the purpose of this article to provide such a demonstration.

In July 1981 the UGC informed the University of Salford that the quota of British students the university was expected to admit was to be reduced by 30 per cent, and that the recurrent grant that the university had been expecting to receive over the next three years would be reduced by some 44 per cent. This precipitated a crisis: the salaries of approximately a third of the staff could no longer be paid from UGC funds, and most of those concerned retired prematurely. Efforts were also made to obtain funds from other sources, and these have been relatively successful;

the university's income is now larger in real terms than it was in 1981, more than half of it coming from sources other than the UGC recurrent grant. What is less well known are the consequences that this restructuring has had on teaching in the university.

Honours degrees awarded by British universities are graded into classes. One measure of the quality of a university is the fraction of its graduates who obtain first-class degrees. Another is the 'average' class of degree, calculated by assigning a value of 5 to a first-class degree; 4 to an upper second; 3 to a lower second; 2 to a third; and 1 to a pass degree and then dividing this sum by the total number of graduates. The first of these measures can be taken as an indication of the extent to which a university acknowledges the exceptionally good student, and the second the extent to which it caters for the average student. Figure 1 (see over) shows how these measures have varied at the University of Salford since 1980; both have shown considerable improvement, but the change in the 'average' class of degree is, statistically, the more significant. In the analysis that follows, I have used these measures of student achievement as indicators of teaching effectiveness; both give the same outcome.

The improvement shown in Fig. 1 is surprising, for the number of academic staff (Table 1, column 1) fell by slightly more in percentage terms than did the number of undergraduate students in the same period (Table 1, column 4). Moreover, this smaller staff was under great pressure to obtain research grant and contract income from non-UGC sources, in which they were very successful; the research income of the university and the

income of the commercial company (Salford University Business Services Ltd), whose profits are covenanted to the university, have both increased considerably (Fig. 2). But how can a smaller number of academic staff, doing more research and consultancy, nevertheless construct an environment in which graduates obtain better classes of degree?

Explanation

The first possible explanation is that the standard of the degrees awarded by the university declined after 1981. This cannot be completely ruled out, but it is highly improbable. All degrees at the university are moderated by external examiners; in any year, about 60 examiners drawn from a wide variety of British universities and industries are involved in our examining boards. A conspiracy to reduce standards steadily over a period of years on the part of so many different professional people (most of whom are strangers to one another) is difficult to imagine.

If the quality of the university's degrees has remained constant, explanations of the changes shown in Fig. 1 must be due either to an increase in the quality of the students admitted or to an improvement in the effectiveness with which those students are taught.

In Britain, the school-leaving examinations (A-levels) are graded A–E; all universities require a minimum of two E-grade passes in two subjects at A-level for admission. Different grades can be given numerical values (A=5, B=4, C=3, D=2, E=1) and thus an 'average' A-level grade of the population of students admitted can be determined. In Fig. 3 the average A-level grade obtained by the entering students is plotted as a function of the year of

Table 1 Changes in the numbers of undergraduate students and graduates (of all kinds) at the University of Salford, 1980–1 to 1987–8

Academic year	1 Full-time UGC-funded staff	2 Full-time graduate students	3 Departmental research assistants	4 Full-time undergraduates	5 Total graduates*	Under-graduates/graduates†
1980–1	469	336	130	4,205	935	4.5
81–2	467	286	123	4,014	876	4.6
82–3	387	288	158	3,731	833	4.5
83–4	348	362	124	3,342	834	4.0
84–5	328	423	169	3,154	920	3.4
85–6	346	517	184	3,104	1,047	3.0
86–7	344	498	201	3,151	1,043	3.0
87–8	335	580	197	3,245	1,112	2.9

* Sum of columns 1, 2 and 3. † Column 4/column 5.

admission. Comparison of Fig. 1 with Fig. 3 shows that there is an approximately three-year lag between the rise in the A-level grades of the students admitted and the rise in the standard of degrees awarded. Because most undergraduate degree courses at Salford last three or four years, this suggests that the increased quality of the output is a result of the increased quality of the input. But closer examination reveals that this cannot account for more than a small fraction of the change. First, a sizeable fraction of Salford students (15–24 per cent, depending on the year) have entering qualifications other than A-levels (usually Business and Technician Education Council awards from colleges of further education). Such students obtain first-class degrees in at least the same proportion as their representation amongst the graduating class.

Second, the correlation between the class of degree awarded to a student and the A-level grades that that student had on entry three or four years previously is 'as poor at Salford during the 1980s as it has been previously², or as it is reported to be by other institutions'. So although the increased academic attainments of most of those admitted to the university undoubtedly contribute to the increase in the quality of the output, the relationship is not simple, and neither can it account for more than a fraction of the changes shown in Fig. 1.

At least as important seems to be the nature of the environment in which the students are taught. In 1982, as part of the process of adaptation that the university had to make to the loss of so much of its UGC support, the senate adopted specific and quite detailed "Aims and Objectives". The section in this document headed "Teaching" states:

The university's teaching is intended to result in graduates . . . with certain . . . characteristics:

(a) the capacity to acquire, organise and systematise knowledge and thereby to develop what the Robbins Committee called 'the general powers of the mind';

(b) the capacity to appreciate, to value and to make judgments of what is beautiful, of good repute and fit for its purpose. This involves the education and training of feelings and emotions as well as implying a moral or ethical framework within which (a) above must be attempted;

(c) the capacity to identify, formulate and then to solve problems and to make, design, organise, produce or construct useful objects and services;

(d) the capacity to co-operate with others, to value communal endeavour and achievement as well as competition.

Although the first two of these objectives are common to most universities, at Salford we believe that the other two are distinctive and that they encapsulate the university's purpose to 'educate for capability'.

This philosophy has helped to shift the

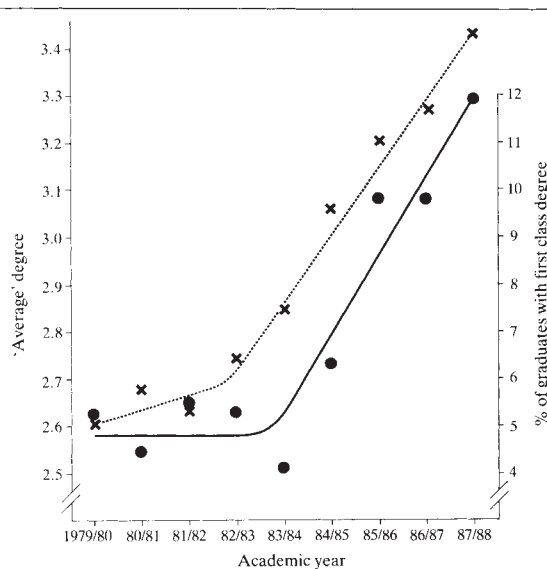


Fig. 1 Degrees awarded by the University of Salford between 1980 and 1988. The 'average' class of degree (crosses) is calculated by assigning a value of 5 to a first-class degree down to 1 for a pass degree, and then dividing the sum by the total number of graduates. The percentage of graduates obtaining a first is plotted by the circles.

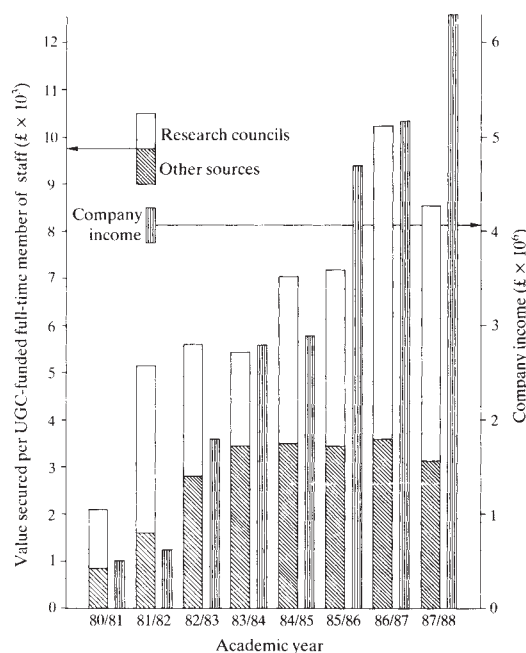


Fig. 2 Research grant and contract income of University of Salford staff, and gross income of the university's wholly owned companies, between 1980 and 1988.

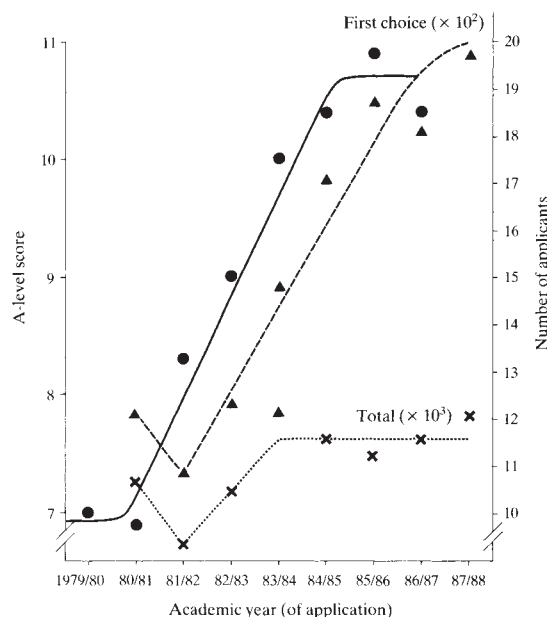


Fig. 3 Characteristics of undergraduate entrants to the University of Salford, and of applications for entry, between 1979 and 1987. Average A-level score of all subjects offered by entrants (circles) is calculated by assigning a value of 5 to an A+ grade down to 1 for an E+ grade, and then dividing the sum by the number of entrants. Total number of applications through the Universities Central Council for Admissions (UCCA) is plotted by the crosses, and number of those who selected Salford as their first choice, of five universities, by the triangles.

emphasis of the university's staff away from a preoccupation with teaching students to a concentration on supporting their capacity to learn — away from an expectation that the university teacher's duty is personally to carry out the teaching, to a realization that the job is much more managerial in nature. Moreover, the success of the university in attracting research contracts (Fig. 2) meant that although the number of academic teaching staff (those financed by UGC) declined, the number of full-time postgraduate students and research assistants paid for out of that income increased. The changes are shown in Table 1. These categories of staff and student have been increasingly involved in teaching.

During the 1980s there has thus been a conscious change in the ethos of the university which has profoundly affected the nature of the teaching offered to undergraduates. The best measure of this change is, I believe, the changing ratio of undergraduates to graduates (of all kinds). The ratio changed, as is shown in Table 1, from 4.5 to 2.9 in the period 1980–1988, but that figure understates the true change; the university now employs many more part-time teachers (many of them former full-time employees who have taken early retirement) and has also been much readier than before to employ industrialists of all kinds in teaching. In 1980–1981 the university spent £46,000 on such part-timers; in 1987–1988 £413,600. The link between teaching and research has thus not been so much individual teacher/researchers as the junior associates whom good researchers are able to employ and who have helped both to change the ethos of the institution and to carry out more of the teaching.

Degree scores

Multiple regression analysis of the degree scores in various departments at different times during the past decade can be used to determine the extent to which the change in A-level grades of the intake and the ratio of undergraduates to graduates have contributed to the changes shown in Fig. 1. Such analyses should not be taken to imply causality or even that the relationship is of the same kind in the many different departments. Among other things, we find:

- In the engineering faculty, 46 per cent of the variance can be accounted for, with the main factor being the ratio of undergraduates to graduates.
- In the science faculty, 55 per cent of the variance can be accounted for, with the main factor being the A-level grades of the intake.
- In the social science and arts faculty, only 13 per cent of the variance can be explained and here, as with engineering, the main factor is the undergraduate to graduate ratio.

The so-called linear subjects, recognizably similar to those taken at school, are found predominantly in the science faculty at Salford, so that these results are not unreasonable.

Over the university as a whole, it is clear that the ratio of undergraduates to graduates has had the main effect (the engineering faculty is the largest at Salford). But nearly half of the variance remains unaccounted for.

One factor which must be important, but which is difficult to quantify, is the motivation of the undergraduates themselves. In Fig. 3, the total number of applications for admission to the university through the Universities Central Council for Admissions (UCCA) has been plotted together with the number of applicants placing Salford first (out of the five choices permitted in the UCCA system). Although the total number of applications has been approximately constant, the number expressing a preference for Salford has increased throughout the period. This factor must be correlated to some extent with motivation and, statistically, may account for some of the unaccounted variance. But it would be naive to conclude that there is a direct causal relationship between degree class and university preference as expressed on the UCCA application form.

Unfortunately, a change in the UCCA system has meant that students are no longer expected to express any preference between their five choices, so this kind of analysis cannot be extended beyond 1988.

No other readily available management statistic seems closely correlated with the changes shown in Fig. 1 although several (such as library use, frequency of borrowing of books and use of computing time, all measured on a *per capita* basis) have increased significantly during this period. A longer run of data, ideally on a departmental or with an individual student basis, will be needed to carry this analysis further and is planned.

The quality of the degrees awarded by the University of Salford has increased markedly during a period of great upheaval, when the number of experienced academic staff was falling both absolutely and in terms of student to staff ratio, and when the demands placed on staff to obtain support from sources other than the UGC for their research (and other) work were intense. This increase in quality is shown both in the performance of those exceptionally good students who obtain first-class degrees and in the 'average' degree awarded. Increases in quality are seen in every faculty and department, and are clearly the result of institutional changes which have affected everyone.

The UGC and other central planning bodies in Britain rarely take into account institutionally based indicators of performance of the kind reported here.

Rather, they consider groupings across disciplines, or 'cost centres', from a national perspective. So such central reviews can rarely detect the changes in institutional ethos which, if the arguments in this paper are accepted, are those most relevant to teaching performance.

Correlation

At Salford, the two changes closely correlated both in timing and extent with the improvements of the quality of the degrees awarded are the improvements of academic performance of entering students and the changing ratio of undergraduate students to graduates of all kinds. The former change is of more significance in the science faculty (departments of mathematics and computing, physics, chemistry and biological sciences) and the latter in the faculties of engineering and social science and arts. Yet for individuals, good A-level grades at entry have been no better predictors of class of degree during the 1980s than at previous times, at Salford and elsewhere^{2,3}.

It is therefore clear that the main factor underlying the improvements shown in Fig. 1 is the change in the ratio of undergraduates to graduates, which in turn is a consequence of the increased effectiveness with which the university's staff have obtained research grants and contracts from research councils and from industrial and other sources. This has been the link between 'good teaching' (as measured in Fig. 1) and research at the University of Salford in the 1980s — together, of course, with the adoption of an explicit philosophy of education ('education for capability') which encouraged and sanctioned the use of the young research assistants and graduate students in the teaching programmes of all departments.

It has been suggested that universities which do not offer the full range of research facilities might be established in Britain¹. This has been interpreted by some to mean that 'teaching-only' universities may evolve. The evidence presented here suggests that such institutions, if established, would be unlikely to produce good graduates as conventionally defined. On the contrary, it would seem that the best way of increasing the teaching effectiveness of a university in Britain is to encourage the staff to do more research and simultaneously to encourage the adoption of a philosophy of education which places more emphasis on student-centred learning and less on staff-centred teaching. □

I am grateful to Dr Rose Baker of Salford University's Academic Information Services for help with the statistical analysis.

1. Advisory Board for the Research Councils *A Strategy for the Science Base* (HMSO, London, 1987).
2. Horlock, J. H. in *The Foundations of the University of Salford* (by Colin Gordon) 207–208 (Univ. Salford, 1975).
3. Bagg, D. G. *Nature* 225, 1105–1108 (1970).

J. M. Ashworth is Vice-Chancellor of the University of Salford, Salford M5 4WT, UK.

Origin of the anomalous ^{40}Ar – ^{39}Ar age of Zaire cubic diamonds: excess ^{40}Ar in pristine mantle fluids

M. Ozima*, S. Zashu*, Y. Takigami* & G. Turner†

* Geophysical Institute, University of Tokyo, Tokyo 113, Japan

† Department of Physics, University of Sheffield, Sheffield, UK

Cubic diamonds from Zaire show excellent correlations between potassium content and $^{40}\text{Ar}/^{36}\text{Ar}$ ratio, and between $^{40}\text{Ar}/^{36}\text{Ar}$ and $^{39}\text{Ar}/^{36}\text{Ar}$, which can be interpreted to yield an 'isochron' age of about 6 Gyr. Now the discovery of a correlation between chlorine content and ^{40}Ar , together with recent mineralogical and geochemical work concerning the origin of cubic diamonds, strongly suggests that the ^{40}Ar is an excess component which has no age significance, and that the ^{40}Ar and its associated potassium are contained in sub-micrometre inclusions of mantle-derived fluid.

ZASHU *et al.*¹ have reported that ten cubic diamonds from Zaire yielded excellent linear correlations in both ^{40}Ar –K and $^{40}\text{Ar}/^{36}\text{Ar}$ versus $\text{K}/^{36}\text{Ar}$ diagrams. If the linear arrays are assumed to be isochrons—which is the basic premise in K–Ar dating—the corresponding ages would be about 6 Gyr, which is older than the age of the Earth. The excellent reproducibility of these results precluded experimental deficiencies; Zashu *et al.* were therefore forced to suggest that either: (1) the ^{40}Ar in the diamonds is excess ^{40}Ar and has no age meaning; (2) the basic assumptions underlying the K–Ar dating method, such as the constancy of the $^{40}\text{K}/\text{K}$ isotope ratio and of the decay constant λ_e of ^{40}K , were not satisfied; or (3) the diamonds or the inclusions in the diamonds are pre-solar. In regard to the second possibility, Akagi and Masuda² and Podosek *et al.*³ analysed Zaire cubic diamonds from the same group used for the K–Ar dating, and found normal K isotope composition. These results also rule out the possibility of a change in λ_e , leaving only the possibilities of excess ^{40}Ar (1) or a pre-solar origin (3). We applied ^{40}Ar – ^{39}Ar step-heating dating to four Zaire cubic diamonds from the same group as those used in the previous K–Ar dating¹ and K isotope studies^{2,3}. The results showed excellent correlation among ^{40}Ar , K and Cl. Correlation of Cl with ^{40}Ar , together with recent mineralogical and geochemical work concerning the origin of cubic diamonds^{4,5}, strongly suggests that the anomalous ^{40}Ar and its associated potassium are contained in sub-micrometre inclusions of mantle-derived fluid.

Samples and experimental details

The samples are cubic diamonds from Zaire and are believed to come from the same group as those in the previous studies^{1–3}. All the samples show a very regular cubic shape of ~4 mm on a side, but their colours differ from dark brown to brownish, yellow and almost white. Two stones from the same batch were cut in half. Cathodoluminescence images of the cut sections revealed an internal zoning consisting of a coat and a core. This structure is quite similar to those reported on several cubic diamonds by Boyd *et al.* (Fig. 1 in ref. 5).

Diamond samples were irradiated by a total neutron flux of $\sim 3 \times 10^{18}$ neutrons per cm^2 (irradiation parameter $J = 1.709 \times 10^{-3}$ for sample 87N01 and $J = 1.91\text{--}1.83 \times 10^{-3}$ for samples 88N01–03) with standard samples MMHb-1(6), K_2SO_4 and CaF_2 . Before the stepped-heating analysis all samples were heated to 200 °C for 6 hours in an attempt to remove adsorbed atmospheric argon. Except for sample 87N01, for which only two temperature fractions were taken, four temperature fractions were taken for each sample: the samples were kept at each temperature step for 30 min for noble-gas extraction. The experimental procedures for noble-gas extraction, purification and

mass-spectrometric analyses are similar to those described previously^{7,8}.

Results

One of the conventional ways of plotting ^{40}Ar – ^{39}Ar data is in the form of 'age spectra'. Because more than 90% of the argon from the diamonds is released in a single temperature step around 1,900 °C, at which point the diamonds are graphitized, there is little to be gained by presenting the data in this form except to note that in most instances the individual ages agree, within error, with the previously reported K–Ar isochron ages of ~6 Gyr. We choose instead to present the more informative isotope-correlation plots⁹. The first of these, Fig. 1, is a conventional isochron plot of $^{40}\text{Ar}/^{36}\text{Ar}$ against $\text{K}/^{36}\text{Ar}$. We have adopted the practice of plotting molar ratios, as this seems the most logical way of comparing elemental ratios with isotope ratios (which are invariably expressed in molar terms). The linear correlation in Fig. 1 suggests that ^{40}Ar is derived from sites occupied by K, implying that it is an indicator of age. The age

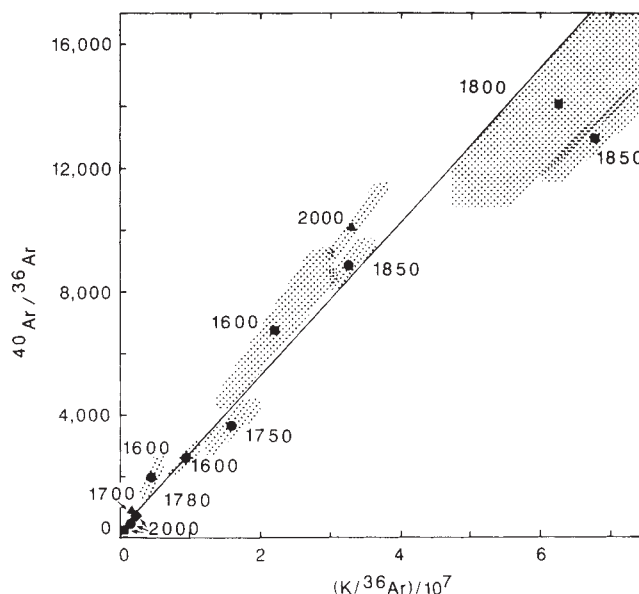


Fig. 1 $^{40}\text{Ar}/^{36}\text{Ar}$ versus $\text{K}/^{36}\text{Ar}$ isochron plot. The ^{40}Ar – ^{39}Ar age inferred from the slope of the line corresponds to 5.7 Gyr. Symbols represent different samples: $\blacktriangle$, 87N01; $\bullet$, 88N01; $\blacksquare$, 88N02; $\blacklozenge$, 88N03. Numerals indicate temperatures (°C). Shaded areas indicate an approximate 1 σ error envelope.

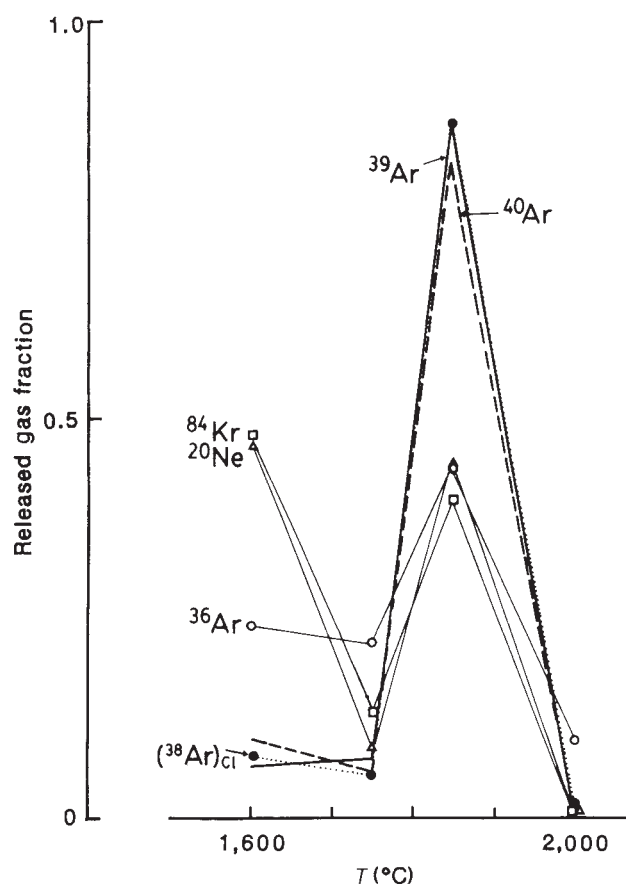


Fig. 2 Thermal-release patterns of noble gases for sample 88N01. $(^{38}\text{Ar})_{\text{Cl}}$ indicates ^{38}Ar corrected for trapped ^{38}Ar (equal to 19% of the ^{36}Ar and corresponding essentially to Cl in the diamonds).

calculated from the slope of the line of best fit is ~ 5.7 Gyr, which is fully consistent with the earlier K–Ar measurements.

In Fig. 2 the fractions of various noble-gas isotopes degassed at each temperature step are plotted against the temperature for sample 88N01. Thermal-release patterns for other samples are almost identical to Fig. 2. The degassing patterns for ^{40}Ar , ^{39}Ar and ^{38}Ar are almost identical, but are distinctly different from those for other noble gases, including ^{36}Ar . This shows conclusively that ^{40}Ar , ^{39}Ar and ^{38}Ar were degassed from similar trapping sites. Because ^{39}Ar is produced by the reaction

$^{39}\text{K}(n, p)^{39}\text{Ar}$, and ^{38}Ar by (n, γ, e^-) reactions of ^{37}Cl , the identical thermal-release patterns for ^{38}Ar , ^{39}Ar and ^{40}Ar indicate that ^{40}Ar was trapped at the same sites as K and Cl. In Table 1 we summarize the analytical data for these ^{40}Ar – ^{39}Ar step-heating experiments.

A two-stage growth model

The excellent correlation among ^{40}Ar , K and Cl can be best explained if these elements existed in the fluid phase before having been trapped in diamonds, as elements in such a fluid could be relatively easily homogenized to give rise to a uniform ^{40}Ar :K:Cl ratio. Because of its incompatible nature, ^{40}Ar can be more easily leached from the country rocks into the fluid phase than can be potassium¹⁰; this would explain the excess ^{40}Ar . If the diamonds trapped such a fluid component to varying degrees, this would give rise to identical thermal-release patterns for ^{40}Ar , K and Cl, and to a linear correlation in the ^{40}Ar –K and ^{40}Ar –Cl diagrams, because they would be trapped as fluid inclusions at the same sites.

Recently, Navon *et al.*⁴ studied several cubic diamonds, including two from Zaire, with ion-microprobe and electron-probe analysis, scanning electron microscopy and infrared spectrometry. They observed that coats in the cubic diamonds contain sub-micrometre inclusions which are characteristically enriched in K_2O (8–22%), CaO (6–18%) and other trace elements. The micro-inclusions are composed of 40–70% metal oxides (SiO_2 , K_2O , FeO and CaO and smaller amounts of Na_2O , MgO , Al_2O_3 , TiO_2 , Cr_2O_3 , P_2O_5 and Cl), 20–50% water and 10–30% CO_2 . Because micro-inclusions were observed commonly in cubic diamonds from various locations, Navon *et al.* suggested that the materials in these micro-inclusions may be related to pristine mantle fluids, which are responsible for widespread metasomatism in the sub-continental mantle.

The difference between the thermal-release patterns of ^{40}Ar , ^{39}Ar , ^{38}Ar and other noble gases (Ne, ^{36}Ar , Kr, Xe) may be explained in terms of a many-stage growth model⁵. Carbon and nitrogen stable-isotope studies suggest that cubic-morphology coated diamonds have undergone two-stage growth: initially a transparent core with octahedral morphology grew in a similar way to common diamonds, and was rimmed by a cubic-habit coat which grew later in contact with the proto-kimerlite host. We suggest that the cores trapped noble gases from the mantle, and that micro-inclusions were incorporated later into the coats during their rapid growth in the presence of fluids. Consequently, the diamonds contain two generations of noble gases: those trapped in the core consist mainly of the normal mantle components, and those contained within micro-inclusions in the coats comprise the excess and *in situ* radiogenic ^{40}Ar .

Table 1 Stepwise thermal degassing of argon isotopes in Zaire cubic diamonds

Sample	<i>T</i> (°C)	^{36}Ar (10^{-9})	^{37}Ar (10^{-12})	^{38}Ar (10^{-9})	^{39}Ar (10^{-12})	^{40}Ar (10^{-6})
87N01 (0.1991 g)	1,700	0.45 ± 0.04	ND*	0.137 ± 0.007	17.3 ± 0.9	0.393 ± 0.009
	2,000	0.30 ± 0.04	ND	0.783 ± 0.008	208 ± 1	3.03 ± 0.09
88N01 (0.2873 g)	1,600	0.121 ± 0.042	3.6 ± 1.8	0.081 ± 0.016	12.8 ± 2.1	0.239 ± 0.02
	1,750	0.038 ± 0.009	2.7 ± 1.3	0.049 ± 0.017	14.2 ± 1.4	0.137 ± 0.01
	1,850	0.225 ± 0.023	26.6 ± 0.4	0.706 ± 0.007	172.2 ± 1.7	1.99 ± 0.02
	2,000	0.051 ± 0.040	1.6 ± 0.4	0.004 ± 0.004	1.7 ± 0.7	0.024 ± 0.001
88N02 (0.2875 g)	1,600	0.036 ± 0.014	5.2 ± 0.6	0.067 ± 0.009	18.3 ± 1.7	0.241 ± 0.01
	1,800	0.143 ± 0.032	70.0 ± 21.0	0.636 ± 0.04	206.7 ± 22.7	2.007 ± 0.05
	1,850	0.228 ± 0.020	2.1 ± 1.3	0.053 ± 0.005	5.7 ± 0.9	0.098 ± 0.005
	2,000	0.027 ± 0.020	~ 0	0.003 ± 0.003	0.17 ± 0.6	0.008 ± 0.011
88N03 (0.3063 g)	1,600	0.136 ± 0.027	1.2 ± 0.5	0.110 ± 0.01	28.6 ± 3.7	0.351 ± 0.018
	1,780	0.098 ± 0.021	0.45 ± 0.28	0.039 ± 0.006	11.2 ± 1.5	0.104 ± 0.008
	1,850	0.192 ± 0.021	10.4 ± 0.7	0.777 ± 0.023	294.6 ± 11.8	2.483 ± 0.050
	2,000	0.075 ± 0.012	~ 0	0.026 ± 0.003	4.0 ± 2.2	0.056 ± 0.005

Units are cm^3 STP g^{-1} .

* The ^{37}Ar peaks were not detected (ND) because of a long time interval (~ 2 yr) after irradiation. However, the correction necessary for these has negligible effect on the apparent age because the ^{40}Ar content is much larger.

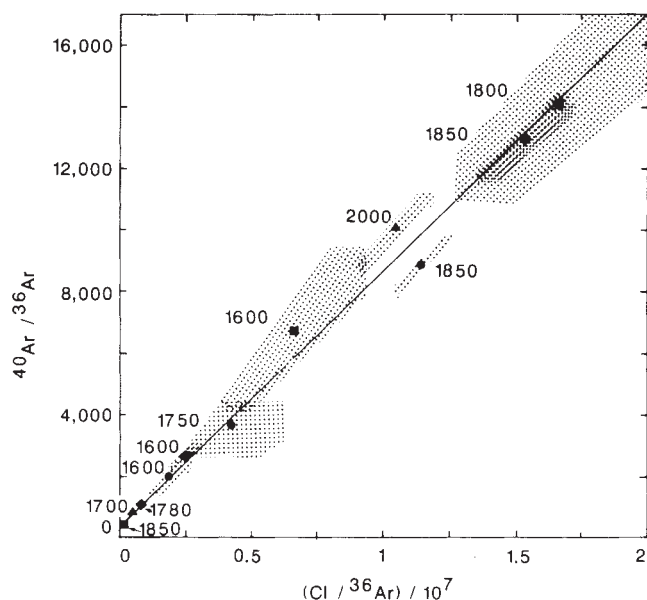


Fig. 3 $^{40}\text{Ar}/^{36}\text{Ar}$ versus $\text{Cl}/^{36}\text{Ar}$ correlation plot. Symbols are the same as those in Fig. 1.

If the cores crystallized at the deeper region in the mantle and were then transported to the upper mantle where they then developed the coat, the two generations of the noble gas would reflect the respective mantle regions. In fact, such a scenario fits quite nicely the suggestions regarding the noble-gas state in the mantle proposed by many authors¹¹⁻¹⁴, whereas the degassed mantle is characterized by a high $^{40}\text{Ar}/^{36}\text{Ar}$ ratio and thereby a high abundance of ^{40}Ar relative to other noble gases, the less degassed deeper mantle retains the primordial noble gases. The argon that is correlated with potassium and chlorine has a very high $^{40}\text{Ar}/^{36}\text{Ar}$ ratio, similar to those commonly observed in mid-ocean-ridge basalts and in contrast to the relatively low $^{40}\text{Ar}/^{36}\text{Ar}$ ratio in the uncorrelated argon (Fig. 2). The latter may correspond to the lower-mantle component. A solar-type neon found in the core of Zaire cubic diamonds⁸ thus would also represent the deeper mantle.

The ^{40}Ar - ^{39}Ar technique can be used to provide additional chemical information in the form of Ca concentrations, by way of the reaction $^{40}\text{Ca}(n, \alpha)^{37}\text{Ar}$, and Cl concentrations, by way of the reaction $^{37}\text{Cl}(n, \gamma, e^-)^{38}\text{Ar}$. Abundances of K, Ca and Cl have been calculated from ^{39}Ar , ^{37}Ar and ^{38}Ar after the usual interference corrections plus a correction for trapped ^{38}Ar (equal to 19% of the measured ^{36}Ar). Bulk element abundances for the four samples analysed, based on a summation of the temperature steps, are given in Table 2. The absolute abundances in the table are probably not particularly significant, in view of the observations by Navon *et al.*⁴ that Ca, K and a number of other elements are concentrated, along with H_2O and CO_2 , in

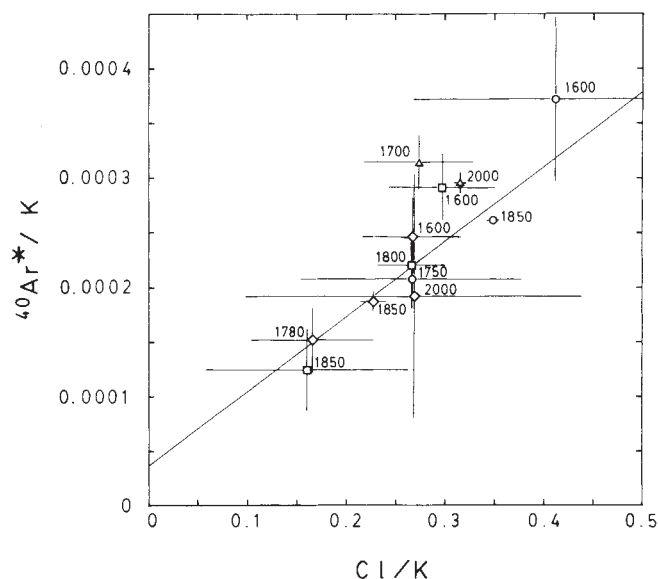


Fig. 4 $^{40}\text{Ar}^*/\text{K}$ versus Cl/K correlation plot. $^{40}\text{Ar}^*$ indicates ^{40}Ar corrected for trapped ^{40}Ar , equal to $340 \times ^{36}\text{Ar}$ (the factor 340 is determined from the y-intercept of the line in Fig. 3). Symbols are the same as those in Fig. 1, but are open rather than filled (for clarity).

sub-micrometre inclusions. If the diamonds that we have analysed contain inclusions similar to those observed by Navon *et al.*, we must conclude that the Cl abundances are also extremely high, possibly as much as 1.2–6 wt%.

In Fig. 3, $^{40}\text{Ar}/^{36}\text{Ar}$ is plotted against $\text{Cl}/^{36}\text{Ar}$. The correlation is excellent and indeed seems superior to the correlation with $\text{K}/^{36}\text{Ar}$. Figure 3 can be interpreted as a mixing diagram between a component with a low $^{40}\text{Ar}/^{36}\text{Ar}$ ratio (equal to 340 ± 80 based on the intercept of Fig. 3) and a chlorine-correlated excess ^{40}Ar component with a molar $^{40}\text{Ar}/\text{Cl}$ ratio of $(8.4 \pm 0.7) \times 10^{-4}$. The situation is analogous to that described in ref. 9 for shallow hydrothermal fluids which have leached Cl and ^{40}Ar from crustal rocks. It is remarkable but not unduly surprising that a similar correlation is found in fluids that we presume to originate in the upper mantle.

Because K/Cl is not constant but shows a factor of two variation (Table 2), $^{40}\text{Ar}/^{36}\text{Ar}$ cannot correlate perfectly with both $\text{K}/^{36}\text{Ar}$ and $\text{Cl}/^{36}\text{Ar}$, and we may therefore investigate whether or not it is possible to separate K-correlated (*in situ* radiogenic?) argon from Cl-correlated (excess?) argon. An attempt to do this is shown in Fig. 4, which is a graph of $^{40}\text{Ar}^*/\text{K}$ against Cl/K . $^{40}\text{Ar}^*$ is the measured ^{40}Ar minus a correction of $340 \times ^{36}\text{Ar}$ for the component represented by the intercept of Fig. 3. Because of the high $^{40}\text{Ar}/^{36}\text{Ar}$ ratios the correction is very small. In spite of the large error bars arising from the small amounts of gas involved, there appears to be a well defined trend in Fig. 4 by which high $^{40}\text{Ar}^*/\text{K}$ ratios (and correspondingly high apparent K-Ar ages) correlate with high Cl/K . The most straightforward interpretation of the correlation is that it represents a mixture of (predominantly) radiogenic argon in a K-rich phase, $\text{Cl}/\text{K} \leq 0.15$, with (predominantly) excess argon in a Cl-rich phase, $\text{Cl}/\text{K} \geq 0.4$. Under this interpretation, the most meaningful K-Ar age estimate is obtained from $^{40}\text{Ar}^*/\text{K}$ for the K-rich endmember. It is not possible to determine this ratio with any accuracy from Fig. 4. The best-fit straight line suggests an age between zero and 3.0 Gyr. Although this age is imprecise, it is comfortably less than the age of the Earth.

The data by Navon *et al.*⁴ gives a C/K ratio (atomic) of ~ 9 . The Zaire cubic diamonds show fairly constant ratios of $\text{K}/^{40}\text{Ar}$ (this study and ref. 1) and $^4\text{He}/^{40}\text{Ar}$ (O.Z. and S.Z., unpublished data), which are 3×10^3 and 0.8 respectively. The $^3\text{He}/^4\text{He}$ ratios in the Zaire cubic diamonds are in the range $(4.2\text{--}13) \times 10^{-6}$ (ref. 8). Hence, we estimate the value of $\text{C}/^3\text{He}$ in the inclusions

Table 2 Argon and element abundances

Sample	^{36}Ar ($10^{-9} \text{ cm}^3 \text{ STP g}^{-1}$)	^{40}Ar ($10^{-9} \text{ cm}^3 \text{ STP g}^{-1}$)	K (p.p.m.)	Cl (p.p.m.)	Ca (p.p.m.)
87N01	0.750	3,423	18.8	5.32	ND
88N01	0.435	2,390	15.0	4.62	5.44
88N02	0.434	2,354	17.4	4.18	12.33
88N03	0.500	2,994	26.0	5.41	1.97

K, Cl and Ca were calculated from ^{39}Ar , ^{38}Ar and ^{37}Ar . The conversion factors for these element abundances are based on measurements of $^{37}\text{Ar}/^{39}\text{Ar}$, $^{38}\text{Ar}/^{39}\text{Ar}$ and the J value (from $^{40}\text{Ar}/^{39}\text{Ar}$) in the monitor MMHb-1, combined with the Ca and Cl abundances determined for MMHb-1 by Roddick¹⁷. Using the notation of Kelley *et al.*¹⁸, we obtain a production rate of Ar from Ca, relative to the rate from K, of $\alpha = 0.472$, and from Cl, relative to K, of $\beta = 12.2$.

to be $\sim 5 \times 10^9$. This is close to the value of 2×10^9 characteristic of the upper mantle^{15,16}, thereby suggesting a common occurrence of the fluids in the mantle.

Conclusions

To summarize, we conclude the following. (1) ^{40}Ar - ^{39}Ar step-heating dating of four cubic diamonds from Zaire gives an apparent age of 5.7 Gyr, which is close to the K-Ar isochron age obtained earlier¹. (2) There is excellent correlation of ^{40}Ar with Cl and K in the cubic diamonds (Figs 1 and 3), and thermal-release patterns for ^{40}Ar , K(^{39}Ar) and Cl(^{38}Ar) are almost identical (Fig. 2). These results show that the ^{40}Ar was trapped with K and Cl, most probably as fluid inclusions. Hence, the extraordinary K-Ar age observed for the Zaire cubic diamonds is attributed to the presence of excess ^{40}Ar . The fluid inclusions are likely to correspond to the micro-inclusions

observed by Navon *et al.*⁴ for several cubic diamonds, including two from Zaire. (3) Most of the other noble gases (Ne, ^{36}Ar , Kr, Xe) reside in different sites to that of ^{40}Ar . These two distinct noble-gases sites can be best explained on the basis of a two-stage growth model for cubic diamonds as proposed by Boyd *et al.*³. This model suggests that whereas most of the ^{36}Ar and other noble gases were trapped in the core of cubic diamonds, ^{40}Ar was trapped later as fluid inclusions in the coat of the diamonds in the upper mantle. The solar-like Ne (ref. 8) and the Ar with low $^{40}\text{Ar}/^{36}\text{Ar}$ ratios, which reside in the core, then represent the mantle region where the core of the diamonds crystallized, which is possibly the deeper mantle.

We thank Drs M. H. Dodson and D. Matthey for helpful reviews. The research was supported by the Japanese Ministry of Education. We are grateful to Mr K. Nishimura for cathodoluminescence analyses.

Received 30 August; accepted 23 November 1988.

1. Zashu, S., Ozima, M. & Nitoh, O. *Nature* **323**, 710-712 (1986).
2. Akagi, T. & Masuda, A. *Geochim. J.* **22**, 81-82 (1988).
3. Podosek, F. A., Pier, J., Nitoh, O., Zashu, S. & Ozima, M. *Nature* **334**, 607-609 (1988).
4. Navon, O., Hutcheon, I. D., Rossman, G. R. & Wasserburg, G. J. *Nature* **335**, 784-789 (1988).
5. Boyd, S. R. *et al.* *Earth planet. Sci. Lett.* **86**, 341-353 (1987).
6. Alexander, E. C. Jr, Michelson, G. M. & Lanphere, M. A. *US geol. Surv. Open File Rep.* **78-701**, 6-8 (1978).
7. Ozima, M. & Zashu, S. *Earth planet. Sci. Lett.* **62**, 24-40 (1983).
8. Ozima, M. & Zashu, S. *Geochim. cosmochim. Acta* **52**, 19-25 (1988).
9. Turner, G. *Geochim. cosmochim. Acta* **52**, 1443-1448 (1988).
10. Hiyagon, H. & Ozima, M. *Geochim. cosmochim. Acta* **50**, 2045-2057 (1986).
11. Hart, R., Dymond, J. & Hogan, L. *Nature* **278**, 156-159 (1979).
12. Kaneoka, I. & Takaoka, N. *Science* **208**, 1366-1368 (1980).
13. Manuel, O. K. & Sabu, D. D. *Geochem. J.* **15**, 245-267 (1981).
14. Allègre, C. J., Staudacher, T., Sarda, P. & Kurz, M. *Nature* **303**, 762-766 (1983).
15. Marty, B. & Jambon, A. *Earth planet. Sci. Lett.* **83**, 16-26 (1987).
16. O'Nions, R. K. & Oxburgh, E. R. *Earth planet. Sci. Lett.* (in the press).
17. Roddick, J. *Geochim. cosmochim. Acta* **47**, 887-898 (1983).
18. Kelley, S., Turner, G., Butterfield, A. W. & Sheppard, T. J. *Earth planet. Sci. Lett.* **79**, 303-318 (1986).

Crystal structure of the active site of ribulose-bisphosphate carboxylase

Inger Andersson*, Stefan Knight*, Gunter Schneider*, Ylva Lindqvist*, Tomas Lundqvist*, Carl-Ivar Brändén* & George H. Lorimer†

* Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala Biomedical Center, PO Box 590, S-75124 Uppsala, Sweden

† Central Research and Development Department, E. I. du Pont de Nemours and Co., Wilmington, Delaware 19898, USA

Ribulose-bisphosphate carboxylase catalyses the key carboxylation reaction of photosynthetic carbon fixation, but also the competing oxygenase reaction of photorespiration. The structure of the active site of ribulose-bisphosphate carboxylase described here should provide a rational basis for attempts to improve the efficiency of the enzyme by genetic engineering.

RUBISCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) is an important target for protein engineering as it controls two competing physiological functions in green plants¹. The enzyme catalyses the carboxylation step in the Calvin cycle of carbon dioxide fixation, a process that stores the energy trapped by photosynthesis. Rubisco also catalyses the initial oxygenation step in photorespiration, during which a considerable amount of the stored energy is converted to heat thereby limiting crop yield. Detailed knowledge of these catalytic mechanisms as well as the structure of the active site is a prerequisite for rational attempts to modify Rubisco by site-directed mutagenesis to increase the carboxylation/oxygenation ratio.

Both catalytic activities of Rubisco require an activation process during which a lysine residue reacts with an activator CO_2 molecule^{2,3}. The labile carbamate formed in this reaction is stabilized by a magnesium ion⁴. The substrate ribulose-1,5-bisphosphate binds to the activated ternary complex and is subsequently either carboxylated by CO_2 or oxygenated by O_2 . All biochemical evidence indicates that these two reactions occur at the same site in the protein¹.

We have previously described⁵ the folding of the polypeptide backbone of non-activated Rubisco from the photosynthetic bacterium *Rhodospirillum rubrum*. This molecule is a dimer of two identical L-chains, each of relative molecular mass (M_r)

55,000 (55 K), which show about 30% amino-acid sequence identity with the L-chains of Rubisco from higher plants⁶. Almost all side chains are now positioned in this model which has been refined at a resolution of 2.3 Å. Here, we describe the active site of this model as well as the binding mode of phosphoglycerate, which is a product in the catalytic reactions.

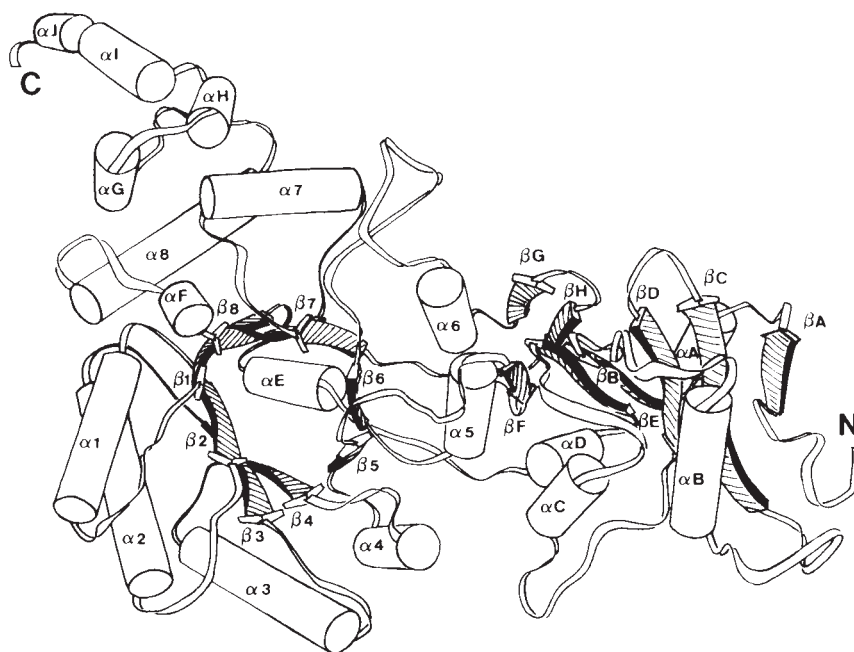
Rubisco molecules from higher plants consist of eight L-chains and in addition eight small polypeptide chains, S, (M_r 15K) of unknown function forming an L_8S_8 enzyme molecule. A model, based on X-ray studies, for the arrangement of the subunits in the non-activated form of the enzyme from tobacco has recently been published⁷. We describe here the X-ray structure of the activated form of Rubisco from spinach with a bound transition state analogue, 2-carboxy-D-arabinitol-1,5-bisphosphate (CABP). In addition, we describe the mode of binding of CABP and the active-site metal atom as well as the effector CO_2 molecule and compare the active sites of the non-activated and activated Rubisco molecules.

Structure determination

Rubisco from *Rh. Rubrum*. X-ray data to 1.7 Å resolution of crystals⁸ of the non-activated form of *Rh. rubrum* Rubisco were collected on 1° oscillation photographs using the wiggler beam line at the synchrotron radiation source in Daresbury, UK. These

Fig. 1 Schematic view of the large subunit of Rubisco from *Rh. rubrum*. β -strands are represented as striped arrows and α -helices as cylinders. The N-terminal domain comprises residues 1–137 and contains β strands β A– β E, which form a mixed five-stranded β -sheet. Helix A is on one side of the β -sheet, and helices B and C are on the other side of the sheet. A region of extended chain after helix D makes the connection to the C-terminal domain. The C-terminal domain has an α/β barrel structure comprising eight parallel β -strands, β 1 to β 8, and eight α -helices, α 1 to α 8 outside the barrel formed by the β -strands. After helix E which closes off the barrel on one side, the chain enters β -strand 1 of the α/β -barrel motif. This domain contains some additional secondary structural features. After helix 6, the polypeptide forms a loop with two antiparallel β -strands (G and H), before the chain enters strand 7 of the α/β barrel. Furthermore, in the loop between β 8 and α 8 there is a short helix (F), which is not part of the barrel motif. After the barrel motif has been completed the polypeptide chain forms helices G–J at the C terminus. These secondary structure elements are assigned to the corresponding amino-acid sequence numbers for Rubisco from *Rh. rubrum* in Table 2.

(Drawing by Ulla Uhlin.)



data were used to refine the initial Rubisco model⁵ by the methods of least-squares^{9,10} and computer graphics modelling¹¹. The crystallographic *R*-factor decreased from 38.6% at 2.9 Å resolution for the starting model to 21.0% at 2.3 Å resolution for the current model.

Native crystals were soaked with 3-phosphoglycerate and a data set was collected to 2.9 Å resolution on a four-circle diffractometer. The positions of bound 3-phosphoglycerate were determined from a difference Fourier map and the structure of the complex has been refined to an *R*-value of 19.6%.

Spinach Rubisco. Crystals of activated enzyme which was cocrystallized with the transition state analogue, CABP, were prepared as previously described¹². Data collection procedures and phasing statistics are presented in Table 1. The L_8S_8 molecule is centered at $x=0$, $y=1/4$ and $z=1/4$ in space group $C222_1$ with a crystallographic twofold axis relating the asymmetric halves of the molecule. The molecule has approximate 422 symmetry as a non-crystallographic fourfold symmetry axis passes through this centre and is inclined at 2.0° to the *c*-axis

in the (101) plane. An envelope of the molecule was defined from an isomorphous, solvent-flattened¹³ map, calculated using heavy atom positions located by Patterson search methods which included the local fourfold symmetry. A final electron density map to 2.8 Å resolution was obtained from phase angles refined by real space averaging¹⁴ of the isomorphous map around the fourfold axis.

A model of the spinach L-chain could easily be built¹¹ into this map, starting from the *Rh. rubrum* Rubisco structure. The map shows clear density for the activator CO_2 as well as for the transition state analogue, CABP. The metal atom position was located from a difference Fourier map of the cobalt-substituted enzyme.

Formation of L_8S_8 Rubisco molecules

The subunit structure of the L-chain in L_8S_8 Rubisco from spinach is fairly similar to that of L_2 Rubisco from *Rh. rubrum*. It is divided into two domains, one smaller N-terminal domain linked to a C-terminal domain which has an eight-stranded α/β barrel structure. The domain arrangement in the subunit and the secondary structure of these domains are illustrated in Fig. 1. The amino-acid residues that form the secondary structure elements in the well defined Rubisco subunit from *Rh. rubrum* are given in Table 2.

Two subunits interact tightly to form the functional L_2 dimeric Rubisco molecule of *Rh. rubrum* (Fig. 2). The core of this binding area consists of interactions between the two C-terminal domains around a local twofold axis. In addition, two regions from the N-terminal domain of one subunit interact with regions from the C-terminal domain of the second subunit. These subunit interactions are of functional importance as some of the residues involved occur in or close to the active-site region.

Each active site of the L_2 dimer is thus built up from residues of both subunits. It is therefore not surprising that similar L_2 dimers occur as part of the L_8S_8 Rubisco molecule from spinach as well as tobacco⁷ and build up the functional active sites. In the spinach enzyme four such dimers are arranged around the local fourfold axis, building up the L_8 core of the molecule. There are large crevices between the ends of the L_2 dimers at the top and the bottom of the L_8 core. The S-subunits occupy these crevices. The overall shape of the molecule resembles a cylinder with a diameter of approximately 110 Å and height 100 Å. The eight active sites are on the outside of the molecule

Fig. 2 A schematic diagram of the L_2 Rubisco dimer illustrating that the core of the dimer interface is formed by interactions between the C-terminal domains of both subunits. In addition the N-terminal domain of one subunit interacts with the C-terminal domain of the second subunit in the vicinity of the active sites. (Drawing by Ulla Uhlin.)

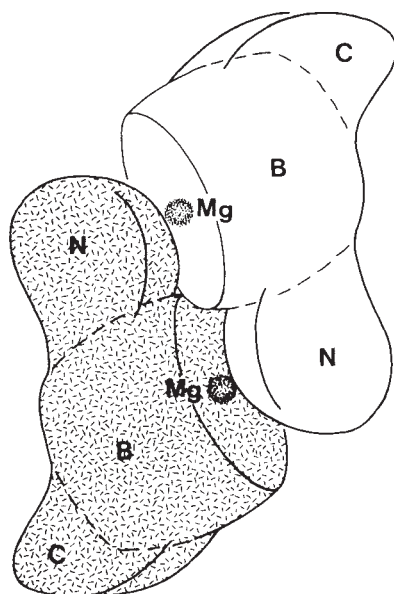


Table 1 Data collection, phasing and map averaging

Mean figure of merit: 0.51 for reflections in resolution range 10–2.8 Å							
Derivative	(dlim (Å))	Wavelength (Å)	X-ray source*	Number of measurements	Number of reflections	$R^{\dagger}$ (10–2.8 Å)	r.m.s. $f_H/E^{\ddagger}$
Native	2.4	0.87	SRS	175.242	70.765	0.049	—
K ₂ Hg(CN) ₄	2.6	1.49	DESY	125.499	54.917	0.051	1.85
KAu(CN) ₂	2.4	0.87	SRS	166.141	68.098	0.069	1.41
EMTS§	2.8	1.69	SRS	58.040	32.117	0.082	1.45
Cob§	2.4	0.87	SRS	140.784	65.836	0.044	—
Map averaging statistics at 2.8 Å resolution							
	R (%)	r.m.s. (degrees)					
Cycle 1	0.41	41					
Cycle 5	0.26	58					
Cycle 9	0.16	63					

Crystals of spinach Rubisco are orthorhombic, space group $C222_1$ with cell dimensions $a = 157.2$ Å, $b = 157.2$ Å and $c = 201.3$ Å. X-ray data were collected on 1° oscillation photographs using synchrotron X-ray sources. Data at 0.87 Å wavelength were obtained from a wiggler beamline. All crystals used were screened for the absence of twinning before use. The data were processed and scaled using the programmes OSC and PROTEIN, kindly provided by J. Remington and W. Steigeman, respectively, and modified by T. A. Jones. Strongly occupied heavy-atom positions were located from difference Patterson maps using an automated search programme written by one of us (S.K.) which searched both Harker and cross-vectors using the non-crystallographic fourfold symmetry of our crystals. The remaining sites were located from difference Fourier maps. For the refinement, 28 positions were used for the Hg derivatives and 12 for the Au derivative. * DESY, The European Molecular Biology Outstation facilities at the Deutsches Elektronen Synchrotron, Hamburg; SRS, Synchrotron Radiation Source in Daresbury, UK. ($R\%$), crystallographic R -factor between observed structure factors and calculated structure factors from the averaged map; r.m.s., phase difference between combined phases and mir phases used in the starting map. $^{\dagger} R = \sum |I - \langle I \rangle| / \sum \langle I \rangle$, $^{\ddagger} f_H$, heavy-atom structure factor; E is the residual lack of closure. § EMTS, ethylmercury thio-salicylate; Cob, native crystals grown in the presence of 5 mM CoCl₂ instead of MgCl₂.

facing the solution. They are spaced widely apart; the shortest distance between the active-site metal atoms is 36 Å. The S-subunits are far from the active sites. There is no contribution to the active site of residues from the S-subunits.

The active site

The active site in the spinach enzyme was identified from the position of the bound transition-state analogue, CABP, as well as the active-site metal atom. The site is shaped like a funnel and is mainly formed by the eight loop regions that connect the eight β -strands with the corresponding helices in the α/β barrel domain. The amino-acid sequences of these loop regions are listed in Table 3 for both *Rh. rubrum* and spinach Rubisco. The N-terminal domain of the second subunit covers part of the top of the active site. In particular, two loop regions of this domain provide residues to the active site. Their sequences are also listed in Table 3.

There is an unusually large number of charged and polar groups in the active site. Within a distance of 10 Å from the

active-site metal ion there are four positively charged (Arg and Lys) and four histidine side chains, four negatively charged (Asp and Glu) and 10 other polar side chains which build up one active site region. All the charged side chains as well as the histidines are conserved in all known species of Rubisco. These residues are boxed in Table 3. All of them except one, Glu 48/60, (the first number refers to the *Rh. rubrum* sequence and the second number to the spinach enzyme) are located in the α/β domain.

The position of the active-site metal atom could be accurately determined from the cobalt-substituted spinach enzyme. It is situated in the middle of the active site close to its bottom and is shielded from solution by CABP. Activator CO₂ at the end of the side chain of carbamylated Lys 191/201 forms one of the protein ligands to the metal (Figs 3 and 4). Substrate CO₂ simulated by the carboxyl group of CABP forms a second ligand. In addition, the side chain of Asp 193/203 as well as the hydroxyl group attached to C2 of CABP (Fig. 4) are close enough to the metal to be ligands. A third invariant residue in loop 2 of the

Table 2 Secondary structure elements and the corresponding amino-acid sequence numbers for the *Rh. rubrum* Rubisco subunit

	Secondary structure		Residue number	Secondary structure		Residue number
N-terminal domain	beta	βA	8–10	alpha	$\alpha 4$	268–278
	alpha	αA	14–20	beta	$\beta 5$	284–287
	beta	βB	23–31	beta	βF	301–303
	alpha	αB	38–49	alpha	$\alpha 5$	305–315
	beta	βC	70–75	beta	$\beta 6$	318–321
	beta	βD	80–86	alpha	$\alpha 6$	336–344
	alpha	αC	101–109	beta	βG	347–349
	beta	βE	120–127	beta	βH	354–356
	alpha	αD	130–133	beta	$\beta 7$	364–369
	alpha	αE	143–150	alpha	$\alpha 7$	376–383
C-terminal domain	beta	$\beta 1$	159–164	beta	$\beta 8$	389–392
	alpha	$\alpha 1$	173–184	alpha	αF	394–398
	beta	$\beta 2$	189–191	alpha	$\alpha 8$	403–419
	alpha	$\alpha 2$	204–222	alpha	αG	423–429
	beta	$\beta 3$	227–231	alpha	αH	431–439
	alpha	$\alpha 3$	237–251	alpha	αI	441–447
	beta	$\beta 4$	258–263	alpha	αJ	451–455

The labelling of the secondary structure elements is as follows: β -strands and α -helices are labelled in alphabetical order starting from the N terminus except those secondary structure elements which build up the α/β barrel in the C-terminal domain. These eight β -strands and α -helices are numbered $\beta 1$ – $\beta 8$ and $\alpha 1$ – $\alpha 8$. The secondary structure elements were assigned using the algorithm of Kabsch and Sander²⁸.

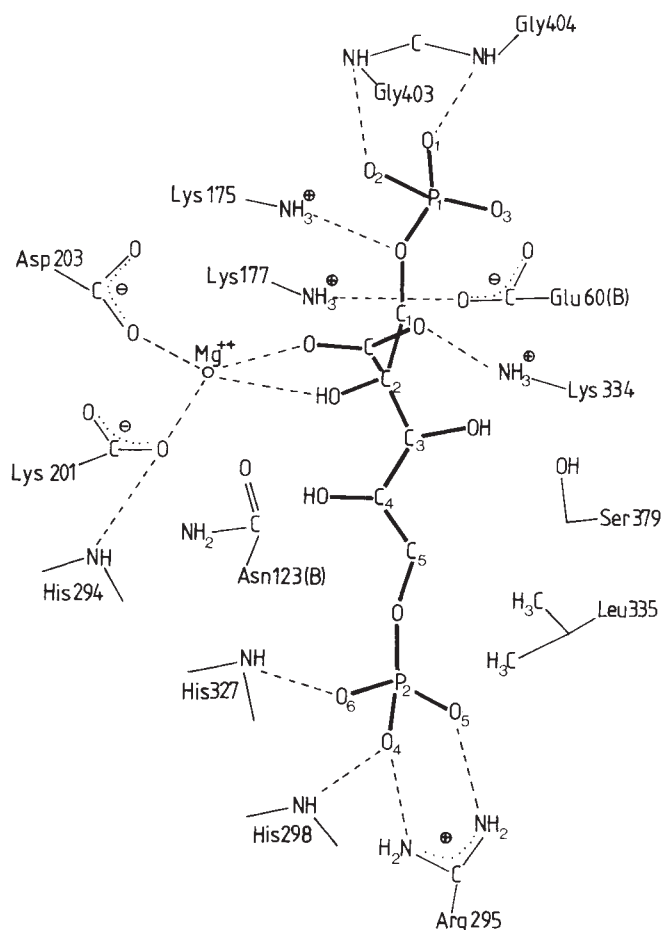


Fig. 3 Schematic diagram of interactions between protein side chains and the metal atom of activated spinach Rubisco as well as a bound transition state analogue, CABP (2-carboxy-D-arabinitol-1,5-bisphosphate).

barrel domain, Glu 194/204, can be model built to ligand the metal through its carboxyl group. The electron density, however, is very weak in our map for this side chain. The invariant residue Asn 111/123 from loop B of the N-terminal domain of the second subunit is also close to the metal atom but too far away (3.5 Å) for strong coordination to the metal atom in our current model.

CABP binds in a rather extended conformation across the barrel (Fig. 4). There are two distinct phosphate-binding sites at opposite sides of the funnel, separating the two phosphorous atoms of CABP by a distance of 9.7 Å. One phosphate-binding site is formed mainly by the adjacent loops 1 and 8 (Table 3). We shall refer to this as phosphate-binding site I. There are two positive charges in the vicinity of the phosphate group, Lys 166/175 and Lys 329/334. In addition two of the phosphate oxygen atoms can form hydrogen bonds to the main chain NH groups of Gly 393/403 and Gly 394/404. These residues are at the N terminus of a short helix in loop 8. Loops 5 and 6 (Table 3) provide the second phosphate binding site II. The side chains of Arg 288/295, His 291/298 and His 321/327 are within 4 Å to oxygen atoms of this phosphate group and surround it with positive charges (Figs 3 and 4).

The sugar bisphosphate molecule is thus anchored at its two ends on opposite sides of the funnel by the phosphate-binding sites and oriented in the middle region by coordination to the active-site metal. In addition the C-4 hydroxyl group interacts with Asn 111/123 from the other subunit (Fig. 4) and the C-3 hydroxyl is close to the invariant residue Ser 368/379. Furthermore, the positively charged groups Lys 166/175 and Lys 329/334 are within interaction distance to the carboxyl group of CABP (Fig. 4) as well as to the phosphate group P1.

We have also studied the binding of the product, phosphoglycerate, to the nonactivated form of the enzyme from *Rh. rubrum*. We find that one molecule of phosphoglycerate binds per active site. When the active site with bound phosphoglycerate is superimposed on the active site of the spinach enzyme with bound CABP we find that the phosphate group of phosphoglycerate binds to phosphate binding site II. Arg 288/295 and His 321/327 provide the positive charges for this binding. The carboxyl group of phosphoglycerate occupies a position close to the position of the metal ion in the activated spinach enzyme. It interacts with the positive side chain of non-carbamylated Lys 191/201. In addition, the asparagine residue from the N-terminal domain, Asn 111/123 is also close to the carboxyl group.

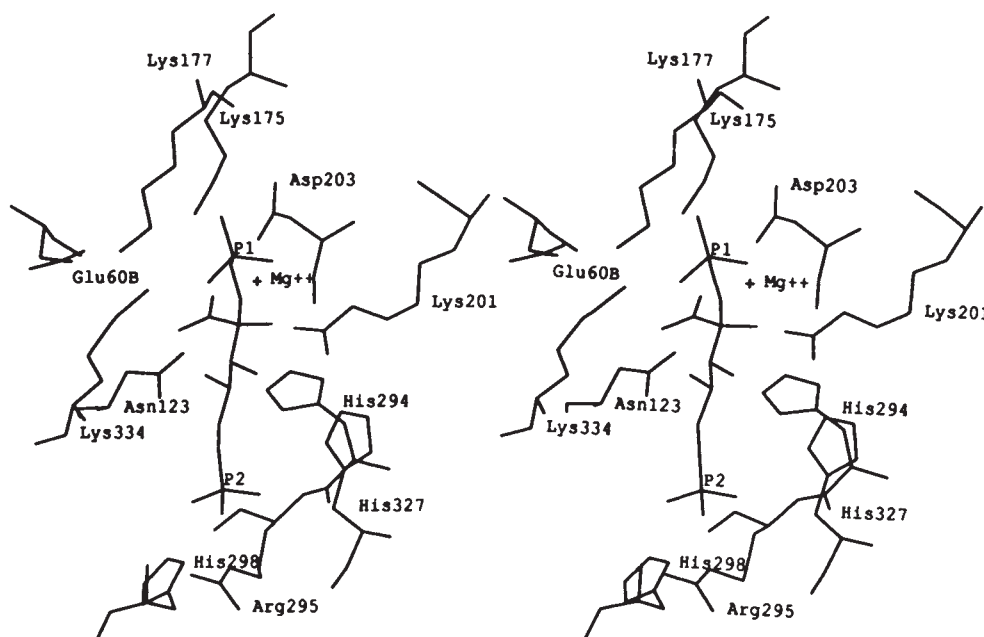


Fig. 4 Computer graphics-generated stereo view of the active site of spinach Rubisco with bound metal atom and CABP.

Table 3 Loop residues which form the active site of Rubisco

C-terminal domain of subunit A

	End of β -strand $\rightarrow$ 1				Loop								1 $\rightarrow$ α helix			
Loop 1																
Loop 2																
Loop 3																
Loop 4																
Loop 5																
Loop 6																
Loop 7																
Loop 8																

N-terminal domain of subunit B

Loop A																
Loop B																

The top two lines of each loop refer to the amino-acid sequence of the spinach enzyme with corresponding residue numbers and the bottom two lines refer to the sequence of the enzyme from *Rh. rubrum*. The numbers of the last residue of the β -strands and the first residue of the α -helices are given. Conserved residues are boxed.

Discussion

The structures described here provide new information on the activation process of Rubisco as well as the identification of a number of specific amino-acid residues in the active site of this enzyme. They also show indirectly that the active-site metal atom is crucial in orienting the substrate molecules for a proper catalytic reaction. Together with biochemical and mutant experiments it can be inferred from this that the metal atom probably also has a more direct role in catalysis by polarizing bonds in reaction intermediates.

Activation is required to form the proper metal-binding site. In the non-activated form Lys 191/201 carries a positive charge and participates in forming a binding site for negatively charged groups as shown in phosphoglycerate binding to the non-activated *Rh. rubrum* enzyme. During the activation process this lysine residue is carbamylated by CO₂ and converted to a negatively charged side chain. The binding site is thereby changed so that it can now accommodate a positively charged metal ion which forms a direct bond to the activator CO₂ group. By binding

magnesium, the active site becomes poised to properly bind and orient the substrate molecules for catalysis. At least in the six-carbon reaction intermediate, for which CABP is an analogue, this is achieved by direct binding, both to one of the hydroxyl groups of the sugar bisphosphate and to the substrate CO₂ group.

This latter conclusion is strongly dependent on correct interpretation of the electron density for CABP. The structural model agrees on this point with electron paramagnetic resonance data^{15,16} which are consistent with metal binding to both these groups. The orientation of CABP in our model also agrees well with the stereochemistry of the catalytic reaction¹⁷ and with NMR measurements on the relative distances between the metal atom and the phosphorous atoms. In our model the distance from the metal to P2 is 7.8 Å and to P1 6.6 Å. The ratio of these distances, 1.2, is in good agreement with NMR measurements¹⁸. Furthermore, by making similar NMR measurements on CABP labelled separately on P1 and P2 with ¹⁷O it has been shown¹⁹ that P1 is closer to the metal atom than P2.

Several mutants of the metal ligands have been constructed using the *Rh. rubrum* gene to study the functional role of magnesium and its ligands both in the overall reaction and in partial reactions²⁰. The results of these studies can now be correlated to the structure of the active site.

To determine whether the charge or the length of the lysine side chain is the important feature of the activation process Lys 191/201 was changed²¹ to Glu. The mutant was completely inactive and could not form a stable complex with CABP. Model building shows that Glu is too short to form a proper metal-binding site instead of carbamylated Lys.

Mutation of the metal ligand Asp 193/203 in *Rh. rubrum* to Asn abolishes the carboxylation reaction²⁰. At high concentrations of magnesium however, the mutant catalyses formation of product when presented with the six-carbon reaction intermediate (2'-carboxy-3-keto-D-arabinitol 1,5-bisphosphate). It did not, however, catalyse an earlier step in the reaction sequence, enolization of the substrate ribulose 1,5-bisphosphate. This early step, which is essential for CO₂ addition and which involves abstraction of a proton from C3 of the substrate thus requires an intact metal-binding site. Enzymatic studies of Rubisco where magnesium has been substituted with different metals have also shown that the nature of the metal strongly influences partitioning between carboxylation and oxygenation²².

Mutants where the adjacent residue Glu 194/204 has been changed to Gln or Val did not catalyse overall carboxylation, nor the partial enolization reaction or product formation from the reaction intermediate²⁰. Clearly this residue has an important function, possibly as a metal ligand.

Received 5 September; accepted 16 December 1988.

1. Mizioro, H. M. & Lorimer, G. H. *Rev. Biochem.* **52**, 507-535 (1983).
2. Lorimer, G. H. & Mizioro, H. M. *Biochemistry* **19**, 5321-5328 (1980).
3. Lorimer, G. H. *Biochemistry* **20**, 1236-1240 (1981).
4. Pierce, J. & Reddy, G. S. *Archs Biochem. Biophys.* **245**, 483-493 (1986).
5. Schneider, G., Lindqvist, Y., Brändén, C.-I. & Lorimer, G. H. *EMBO J.* **5**, 3409-3415 (1986).
6. Nargang, F., McIntosh, L. & Somerville, C. R. *Molec. Gen. Genet.* **193**, 220-224 (1984).
7. Chapman, M. S., Suh, S. W., Cascio, D., Smith, W. W. & Eisenberg, D. *Nature* **329**, 354-356 (1987).
8. Schneider, G., Brändén, C.-I. & Lorimer, G. H. *J. molec. Biol.* **187**, 141-143 (1986).
9. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **A33**, 800-804 (1977).
10. Konnert, J. H. & Hendrickson, W. A. *Acta crystallogr.* **A36**, 344-350 (1980).
11. Jones, T. A. & Thirup, S. *EMBO J.* **5**, 819-822 (1986).
12. Andersson, I. & Brändén, C.-I. *J. molec. Biol.* **172**, 363-366 (1984).
13. Wang, B. C. *Meth. Enzym.* **115**, 252-270 (1985).

Identification of the base responsible for the proton abstraction in the enolization reaction has prompted a number of mutant as well as biochemical experiments. Modification of Lys 166/175 by affinity labelling²³ as well as mutations²⁴ abolishes or drastically decreases catalysis of the enolization reaction. It has therefore been concluded^{25,26} that Lys 166/175 is the base responsible for proton abstraction in this partial reaction. In our spinach Rubisco model, which simulates a later stage in the reaction sequence, the nitrogen atom of this side chain is not close to C3 of CABP but instead interacts with the P1 phosphate group and the carboxyl group (Fig. 4). Model building outside the electron density could bring the nitrogen atom within 3 Å from C3. There are, however, four other side chains in our model which are close to C3 of CABP and are also possible candidates for proton abstraction; His 321/327, Lys 329/334, Ser 368/379 and the nitrogen atom of the carbamate group at Lys 191/201.

Glu 48/60 is part of an invariant sequence in loop A of the N-terminal domain. In both structures this residue is involved in subunit interactions with Lys 168/177 from the C-terminal domain of the other subunit (Figs 3 and 4). Mutation of this residue abolishes or drastically reduces activity²⁷, presumably due to changes in the subunit interactions which could influence the position of Asn 111/123 from loop B of the N-terminal domain which is close to the metal. The active site would no longer be intact and catalysis would be affected.

We thank the staff at the Daresbury synchrotron radiation facility, UK for assistance with the data collection. This work was supported by grants from the Swedish research councils NFR and SJFR.

14. Bricogne, G. *Acta crystallogr.* **A32**, 832-847 (1976).
15. Mizioro, H. M. & Sealy, R. C. *Biochemistry* **23**, 479-485 (1984).
16. Styring, S. & Brändén, R. *Biochemistry* **24**, 6011-6019 (1985).
17. Schloss, J. V. & Lorimer, G. H. *J. biol. Chem.* **257**, 4691-4694 (1982).
18. Pierce, J. *Biochemistry* **27**, 1311-1315 (1988).
19. Lorimer, G. H., Gutteridge, S. & Reddy, G. S. *J. biol. Chem.* (in the press).
20. Lorimer, G. H., Gutteridge, S. & Madden, M. W. in *Plant Molecular Biology* (eds Wettstein, D. & Chua, N.-H.) 21-31 (Nato ASI series A, 140).
21. Estelle, M., Hanks, J., McIntosh, L. & Somerville, C. R. *J. biol. Chem.* **260**, 9523-9526 (1985).
22. Robison, P. D., Martin, N. N. & Tabita, F. R. *Biochemistry* **18**, 4453-4458 (1979).
23. Hartman, F. C., Stringer, C. D. & Lee, E. H. *Archs Biochem. Biophys.* **232**, 280-295 (1984).
24. Hartman, F. C. *et al. J. biol. Chem.* **262**, 3496-3501 (1987).
25. Herndon, C. S., Norton, I. L. & Hartman, F. C. *Biochemistry* **21**, 1380-1385 (1982).
26. Lorimer, G. H. & Hartman, F. C. *J. biol. Chem.* **263**, 6468-6471 (1988).
27. Larimer, F. W., Lee, E. H., Mural, R. J., Soper, T. S. & Hartman, F. C. *J. biol. Chem.* **262**, 15327-15329 (1987).
28. Kabsch, W. & Sander, C. *FEBS Lett.* **155**, 179-182 (1983).

LETTERS TO NATURE

Discovery of a double radio source associated with Cygnus X-3

R. G. Strom*, J. van Paradijs†‡ & M. van der Klis§

* Netherlands Foundation for Radio Astronomy, PO Box 2, 7990 AA Dwingeloo, The Netherlands

† Astronomical Institute "Anton Pannekoek", University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

‡ Center for High-Energy Astrophysics, Nikhef-H, Kruislaan 409, 1098 SL Amsterdam, The Netherlands

§ EXOSAT Observatory, Space Science Department of the European Space Agency, ESTEC, PO Box 299, 2200 AG Noordwijk, The Netherlands

The combination of high X-ray luminosity¹ ($\sim 10^{38}$ erg s⁻¹), short orbital period² (4.8 hours), large non-thermal radio outbursts³⁻⁵ and reports of very-high-energy (up to 10^{15} eV) γ -ray emission^{6,7} (see also ref. 8) and muon-rich showers⁹⁻¹¹ makes Cyg X-3 a unique object among X-ray binaries. Very-high-resolution spatial mapping shows that the radio source starts expanding at the beginning of outbursts and becomes elongated^{4,5,12}, indicating that Cyg X-3 contains a relativistic jet, similar to SS433 (ref. 13). We report

here the discovery of a weak (~ 4.5 mJy), arcsecond-scale double radio source located symmetrically with respect to the compact variable radio core of Cyg X-3. The double radio source has a separation of about 6 arcsec and is roughly aligned with the elongation direction of the sub-arcsec core from Very-Long-Baseline Interferometry observations^{12,14,15}. Like the latter it has a rather broad opening angle. This suggests a generic connection between the inner and outer radio structures, and implies that the (poor) collimation evident in the large-scale double source originates near the binary system.

The observations described here were carried out on 14 October 1987 using the Westerbork Synthesis Radio Telescope at 4,874 MHz (6 cm) with an 80-MHz bandwidth, providing an r.m.s. noise level of 0.08 mJy after 12 hours. We did the measurements in standard synthesis mode¹⁶ as part of a long-term study of the variability of the source. In view of this variability (see Fig. 1), we have taken special steps to obtain a map with the required dynamic range. Standard redundancy self-calibration¹⁷ and phase alignment using the unresolved central source enabled us to obtain a dynamic range in the final map (referred to the strength of the central source) in excess of 30 dB. To remove the central component and its near-in sidelobes, we have subtracted a point source from the map at the known position of Cyg X-3 using the flux density given by the measured light curve.

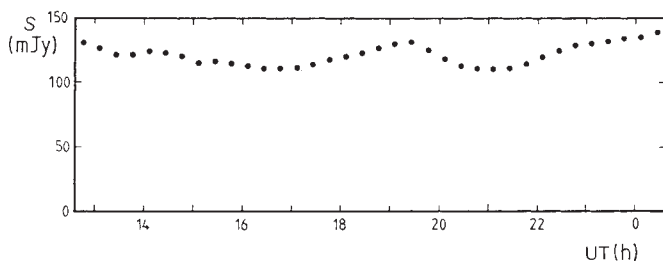


Fig. 1 The 6-cm flux density of the central core of Cygnus X-3 shown as a function of time (UT) on 14 October 1987. Each data point is an average over 20 minutes.

Because our angular resolution is only just sufficient to separate the outer emission from the intense core, there is some uncertainty in the exact level of the central extended emission which underlies the core, and also in the total flux density quoted below. Two other long observations from the present series, of 11 and 9 hours, reproduce the faint double-lobe structure and provide an independent check upon the reliability of our map. The extended structure surrounding Cyg X-3 is shown in Fig. 2. The dominant central component (indicated by a cross) straddles a symmetrical double source with a separation of ~ 6 arcsec. The axis of this double structure is elongated approximately north-south. The opening angle of the components, when seen from the central source, is about 75° . Despite its low flux density (4.5 mJy in total, split roughly equally between the two components), there is little doubt that the outer double is part of Cyg X-3; its morphology and near-perfect alignment with the central component, and the north-south elongation of the central component on sub-arcsec scales^{12,14,15}, make the association virtually certain.

For a distance of 12 kpc (ref. 1) we can use the observed angular size, component dimensions and flux densities to derive physical parameters for the Cyg X-3 outer double. To compute the source volume, each component has been approximated by a cone of base diameter 0.4 pc and apex height 0.2 pc. Because of their near-equal flux densities and sizes, we take the components to be alike in all respects and give values only for the entire source. We do not know the spectral index α (defined by the flux density (S)-frequency (ν) relationship $S_\nu \propto \nu^\alpha$) of the double source, but we will assume that the spectrum is non-thermal and hence that the emission is incoherent synchrotron radiation. If we take for α values in the range from -0.5 to -1.1 , as observed for the lobes of Sco X-1 (ref. 18) and similar

emission, which contrasts strongly with Sco X-1 (ref. 20). If the radio lobes are the result of beamed energy transport from the central binary system, then their breadth could be ascribed either to a beam that often changes direction (as is actually observed in the compact radio component of SS433 (ref. 13)) or to one that is intrinsically broad. Observational evidence from Very-Long-Baseline Interferometry (VLBI) investigations is somewhat ambiguous, however. Recent measurements¹² of a low-level flare suggest that the inner sub-arcsec structure, which is directed north-south, has a very large opening angle of $\sim 80^\circ$. Although observed on a scale some 5,000 times smaller than the outer components, the VLBI structure subtends essentially the same angle (as indicated schematically in Fig. 2). This would appear to suggest that collimation rather than beam wobble is the cause of the outer components' breadth. VLBI maps made during major outbursts, on the other hand, favour the alternative point of view. Studies of flares observed in October 1983 reveal well-collimated double structures, one¹⁴ elongated in position angle 33° (within but not far from the boundary of the opening angle shown in Fig. 2), and another¹⁵ directed north-south. Thus, although there is evidence for a low degree of collimation during some flare events (to the present time, those of low intensity), beam wobble is observed in well-collimated (and thus far major) outbursts. What is clear is that all of the VLBI structures observed so far are contained within the opening angle defined by the outer components shown in Fig. 2.

In Cyg X-3, as in Sco X-1 and SS433, most of the power is radiated at high (optical and X-ray) frequencies by predominantly thermal emission processes close (10^{11} - 10^{12} cm) to the binary system. Although the radio emission is energetically insignificant, only the VLBI components (on a scale of 10^{14} - 10^{15} cm) exhibit a morphological link to the outer radio lobes, and we argue that the two structures can be tied together on the basis of a twin-jet model, as has been done for SS433 (ref. 13). We will use the observed and inferred properties of the inner and outer radio emission to investigate the timescale required to generate the outer lobes. The 'quiescent' (slowly varying) emission from the centre of Cyg X-3 has a typical flux density of 200 mJy at frequencies ranging up to 20 GHz (ref. 21). About twice a year²² there are major (10 Jy or more) outbursts lasting for a few days. The total power output from both

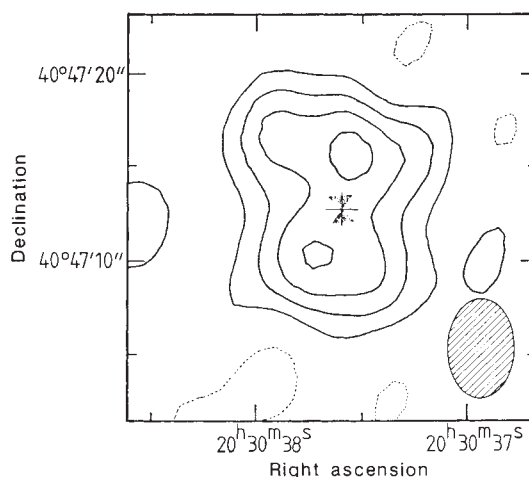


Fig. 2 The extended emission surrounding Cygnus X-3 observed at 6 cm. Contours have been drawn at ± 0.21 , 0.42, 0.63 and 0.84 mJy per beam. The dominant central component (its position indicated by a cross) has been removed from the map using the flux densities shown in Fig. 1. The hour-glass-shaped shaded region in the centre shows schematically the opening angle of the VLBI components whose true scale is some 5,000 times smaller than the extended structure. The $3.5'' \times 5.3''$ half-power beam is shown hatched in the lower right-hand corner.

Table 1 Observed and derived radio-source properties

	Cyg X-3	Sco X-1	SS433
Distance (kpc)	12	0.5	5
S_6 , extended emission (mJy)	4.5*	17	100
Size, extended emission (pc)	0.35	0.38	0.12
$E_{\min}$, outer components (erg)	10^{45}	8×10^{40}	10^{43}
B_{eq} , outer components (G)	3×10^{-4}	5×10^{-3}	10^{-3}

* Based on assumed flux densities of 7.6 Jy for 3C147 and of 7.4 Jy for 3C286.

to values found in extragalactic non-thermal radio sources, we find that the minimum energy¹⁹ varies from 0.6×10^{45} to 1.7×10^{45} erg. The parameter values are summarized in the first column of Table 1 (where we assume $\alpha = -0.8$). The extended structures of Cyg X-3 and Sco X-1 are similar not only in their double morphology but also in overall linear size, although the other physical parameters are quite different (Table 1).

Apart from its double structure, the most striking morphological feature of Cyg X-3 is the large lateral extent of the outer

processes (in which the former dominates), P , is $\sim 3 \times 10^{33} \text{ erg s}^{-1}$. If a similar amount of power drives the outer lobes then the timescale implied is $E_{\text{min}}/P = 11,000 \text{ yr}$ (where radiation losses have been ignored, a point we consider below). The corresponding average recession speed of the lobes is nearly 20 km s^{-1} . Although there is no independent evidence as to how energy is shared between the inner and outer radio components, we note that this average speed inferred for Cyg X-3 is similar to that actually measured in one of the lobes of Sco X-1 (ref. 20). The speed at which radio-emitting particles reach the outer lobes may be much higher than this, as implied by the expansion rates of $0.2\text{--}0.4c$ measured using VLBI¹⁴, and by the fact that the components produced by major flares achieve separations of 0.1 arcmin or more after several weeks^{14,15}.

Radiation losses, which we have ignored in the above estimates, are unlikely to be important provided that the magnetic field strength does not greatly exceed its equipartition value. From standard formulae¹⁹ the lifetime of a synchrotron electron emitting 5-GHz radiation in the equipartition field (Table 1) is $1.4 \times 10^5 \text{ yr}$. This is longer than the age found above, so radiation losses probably do not play a significant role. They do, however, put an upper limit on the age, which could be refined by future determinations of the radio spectrum. Moreover, if the jet efficiency has not changed dramatically with time, Cyg X-3 cannot have been much less active in the past if the outer-component age is to be less than the synchrotron lifetime.

Six of the ten brightest low-mass X-ray binaries (LMXBs) in the Fourth Uhuru Catalogue²³ have been detected as weak radio sources²⁴. On the other hand, in a survey of 21 LMXBs by Grindlay and Seaquist²⁵, only three were detected. These results suggest that radio emission is relatively common in LMXBs but that the radio luminosities vary substantially, a conclusion supported by detailed observations of Sco X-1 (ref. 26), GX13+1 (ref. 27) and GX17+2 (ref. 28). Furthermore, three of the radio-emitting X-ray binaries, Cyg X-3, Sco X-1 and SS433, are now known to have produced extended double radio structures; SS433 also exhibits double-lobed X-ray emission²⁹. It is possible that such structures are the rule rather than the exception in LMXB systems, but that they are usually not observed because they are too weak.

The double structure associated with binary systems contrasts strikingly with the morphologies produced and powered by isolated neutron stars. There is no doubt that both the Crab Nebula³⁰ and the flat spectrum component in CTB80 (refs 31, 32) are maintained by their associated neutron stars/pulsars PSR0531+21 and PSR1951+32, even though their rates of rotational energy loss differ by orders of magnitude. Neither pulsar is in a binary system and both have produced rather amorphous but somewhat elongated synchrotron nebulae. Their structures are quite different from those observed in Sco X-1 and Cyg X-3, a difference which can probably be ascribed to the presence of an accretion disk in the binary systems. Similar arguments are likely to apply to galactic nuclei powering extragalactic double radio sources.

Finally one might wonder why this extended emission was not detected earlier, given that Cyg X-3 has been the subject of intensive scrutiny for over 15 years. One possibility is that, as in Sco X-1 (ref. 20), the outer components vary in intensity. It seems more likely that the extended lobes were not detected previously for instrumental reasons, the major impediment undoubtedly being the strength and variability of the dominant core component. The excellent dynamic range of the Westerbork Telescope was an essential element in our discovery, although the requirements of the present detection (somewhat better than 200:1) are rather modest by the standards of modern synthesis arrays¹⁷. Our success may derive largely from an angular resolution that happens to be well-matched to the outer lobe dimensions. Unless a significant fraction of the emission in Fig. 2 emanates from sub-arcsecond features, an observation with, say, a 1-arcsec beam would have to reach a noise level of 0.02 mJy

and a dynamic range of at least 6,000:1 to obtain a reliable detection. With greater resolution, or when the core is more active or at shorter wavelengths, the demands will be even higher. Variable cores may well be masking similar extended structures in other binary systems.

The Westerbork Synthesis Radio Telescope is operated by the Foundation for Radio Astronomy with financial support from the Netherlands Organisation for Scientific Research (NWO). We are grateful to Dr R. M. Hjellming for several useful suggestions.

Note added in proof: Further analysis of our data suggests that the extended emission is (partially) polarized, establishing its non-thermal nature.

Received 9 October; accepted 7 December 1988.

1. Dickey, J. M. *Astrophys. J.* **273**, L71-L73 (1983).
2. Brinkman, A. et al. *IAU Circ.* no. 2446 (1972).
3. Gregory, P. C. et al. *Nature* **239**, 440-443 (1972).
4. Geldzahler, B. J. et al. *Astrophys. J.* **273**, L65-L69 (1983).
5. Spencer, R. E., Swinney, R. W., Johnston, K. J. & Hjellming, R. M. *Astrophys. J.* **309**, 694-699 (1986).
6. Samorski, M. & Stamm, W. *Astrophys. J.* **268**, L17-L21 (1983).
7. Lloyd-Evans, J. et al. *Nature* **305**, 784-787 (1983).
8. Hermsen, W. et al. *Astr. Astrophys.* **175**, 141-150 (1987).
9. Marshak, M. L. et al. *Phys. Rev. Lett.* **54**, 2079-2082 (1985).
10. Battistini, G. et al. *Phys. Lett. B* **155**, 465-467 (1985).
11. Bartelt, J. et al. *Phys. Rev. D* **32**, 1630-1641 (1985).
12. Molnar, L. A., Reid, M. J. & Grindlay, J. E. *Astrophys. J.* **331**, 494-508 (1988).
13. Hjellming, R. M. & Johnston, K. J. *Astrophys. J.* **246**, L141-L145 (1981).
14. Hjellming, R. M. in *Galactic and Extragalactic Radio Astronomy* (eds Verschuur, G. L. & Kellermann, K. I.) 381-438 (Springer Astronomy & Astrophysics Library, New York, 1988).
15. Spencer, R. E. & Johnston, K. J. in *RS Ophiuci (1985) and the recurrent Nova Phenomenon* (ed. Bode, M. F.) 215-219 (VNU Science, Utrecht, 1987).
16. Bos, A., Raimond, E. & van Someren Gréve, H. W. *Astr. Astrophys.* **98**, 251-259 (1981).
17. Noordam, J. E. & de Bruyn, A. G. *Nature* **299**, 597-600 (1982).
18. Geldzahler, B. J., Fomalont, E. B., Hildrup, K. & Corey, B. E. *Astr. J.* **86**, 1036-1041 (1981).
19. Pacholczyk, A. G. *Radio Astrophysics* (Freeman, San Francisco, 1970).
20. Geldzahler, B. J. & Fomalont, E. B. *Astrophys. J.* **311**, 805-813 (1986).
21. Molnar, L. A., Reid, M. J. & Grindlay, J. E. *Nature* **310**, 662-665 (1984).
22. Johnston, K. J. et al. *Astrophys. J.* **309**, 707-713 (1986).
23. Forman, W. et al. *Astrophys. J. Suppl.* **38**, 357-412 (1978).
24. Hjellming, R. M. & Johnston, K. J. in *The Physics of Accretion onto Compact Objects* (eds Mason, K. O., Watson, M. G. & White, N. E.) Springer Lecture Notes in Physics **266**, 287-301 (1986).
25. Grindlay, J. E. & Seaquist, E. R. *Astrophys. J.* **310**, 172-175 (1986).
26. Priedhorsky, W. et al. *Astrophys. J.* **306**, L91-L95 (1986).
27. Garcia, M. R. et al. *Astrophys. J.* **328**, 552-558 (1988).
28. Penninx, W. et al. *Nature* **336**, 146-148 (1988).
29. Watson, M. G., Willingale, R., Grindlay, J. E. & Seward, F. D. *Astrophys. J.* **273**, 688-696 (1983).
30. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
31. Strom, R. G. *Astrophys. J.* **319**, L103-L107 (1987).
32. Kulkarni, S. R. *Nature* **331**, 50-53 (1988).

Is PSR1957+20 eclipsed by a comet, magnetosphere or particulate cloud?

F. Curtis Michel

Space Physics and Astronomy, Rice University, PO Box 1892, Houston, Texas 77251, USA

The extraordinary eclipsing binary pulsar PSR1957+20 is in eclipse over eight per cent of its orbit¹. This is more than twice the extent of the companion's Roche lobe, the maximum stable synchronous configuration. A possible explanation² for the large eclipse is that it is caused by a plasma wind driven from the companion and forming a comet-like tail³. But an ionized cometary tail implies a refractive eclipse, which would produce large variations in pulse arrival times; this is inconsistent with the small time delays observed. Here we suggest two alternative explanations. One possibility is that the companion has a typical white-dwarf magnetic field ($\sim 10^5 \text{ gauss}$), and that the eclipse is then due to the plasma-filled magnetosphere which is confined and blown back sharply by the pulsar wind, with the time delays produced by a much less dense tail. In the other limit of negligible magnetization of the companion, an eclipse could be caused by a particulate cloud (comprising particles of size $\sim 10 \text{ cm}$) which should form

about the companion. We assume that the companion is the remnant of an evolved companion star that is now being excited by the 622-Hz pulsar radiation. Such excitation could either supply plasma to the magnetosphere (for the magnetized case) or pull off neutral gas and deposit it in the vicinity to sustain a particulate cloud (for the non-magnetized case). Near-infrared radiation from this variable system might test the cloud hypothesis, whereas time delays in a magnetized tail suggest radio-frequency polarization tests for the magnetosphere model.

PSR1957+20 consists of a 1.6-ms pulsar and a $0.02 M_{\odot}$ companion¹. The eclipse duration (8×10^{10} cm of orbital trajectory for a pulsar mass of $1.4 M_{\odot}$ and therefore a separation $a = 1.7 \times 10^{11}$ cm), is more than a factor of two larger than the Roche-lobe distance (L_1 is located at 2.8×10^{10} cm, and the equipotential surface from that point extends $\pm 1.9 \times 10^{10}$ cm along the orbital axis), so the companion apparently cannot be physically large enough to block the pulsar signal (for this mass, the companion's radius is presumably $\sim 3 \times 10^9$ cm (ref. 4)).

The eclipses of PSR1957+20 by its companion seem quite diagnostic, and seem to be real rather than indicating loss of lock with the pulsar owing to increasing time delay (D. Stinebring, personal communication). For zero orbital eccentricity and an eclipse extending from 0.21 to 0.29 of phase (zero phase is the ascending node and unit phase corresponds to a full revolution), to first order, the eclipse is symmetric and centred on inferior conjunction (that is, as if the pulsar were occulted by the body of an orbiting sphere 4×10^{10} cm in radius). This degree of symmetry rules out almost immediately eclipses of the pulsar by a cometary structure around the companion, because, owing to the intrinsic curvature of such cometary tails, the flanks would be intercepted at different distances and phase angles on entry and exit from eclipse. This is indicated in Fig. 1, where we have numerically simulated the cometary shape; a sharp-entry eclipse would obtain, but not a sharp-exit eclipse. This confirms the sketch given in ref. 3.

In the simulation, we assumed a standard outflow wind of 600 km s^{-1} and radial wind acceleration 12 times the local pulsar gravitational acceleration (chosen only to give a roughly correct cross-sectional size). This model is a direct generalization of a standard one for comets, in which, because of the smaller relative scales involved, the solar-wind acceleration is often approximated as uniform and unidirectional. Of course, this refers to individual particle trajectories, but the overall morphology conforms to the obvious expectations. One can imagine effects that would cause the cometary tail to flare more (plasma pressure gradients within the tail, for example, which must be outward), but not less. E. S. Phinney (personal communication) has suggested that the tail might be highly collimated, expanding only to a fixed cross-section, which would lead to shadowing of the tail by the cometary head; however, shadowing of a relativistic wind would surround the tail with a vacuum region, and rather specialized properties would then be required of the tail plasma for it not to expand to fill these regions, that is, for it not be collimated in the first place. More realistic calculations are clearly needed.

There is a measurable time delay of up to at least 400 μs , which trails off over another 15° following the eclipse, but there is essentially no time delay preceding eclipse. Although it is tempting to suppose that this time delay is directly related to propagation through the tail in Fig. 1, it is too small in magnitude to give a sharp eclipse by having this plasma refract the radio pulses away from the observer. Refraction by plasma occurs away from high plasma concentrations and hence permits eclipses, whereas refraction by neutral gas is towards high gas concentration and hence does not completely interrupt the signal; the Voyager spacecraft was tracked continuously behind Uranus, for example (V. R. Eshleman, personal communication), because of the latter effect. Given that the radius of the system is ~ 5.8 light-seconds, the average index of refraction along a line of sight that reaches to within 4×10^{10} cm of the

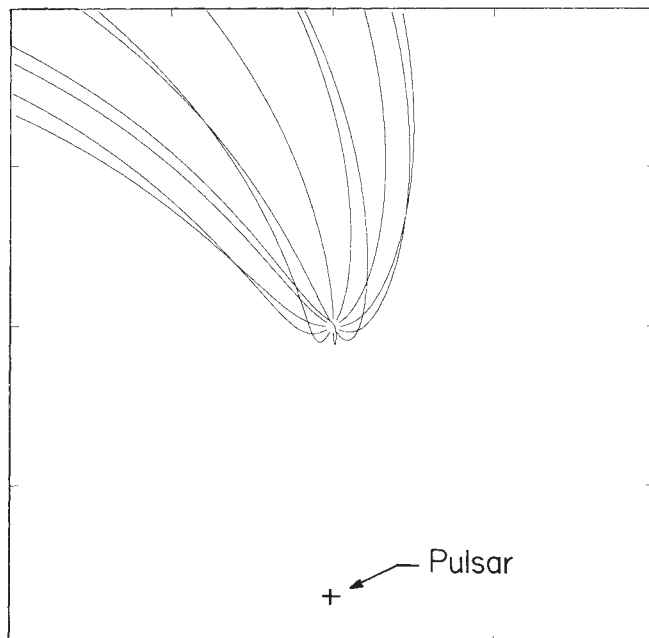


Fig. 1. Simulated cometary wind. Individual particle trajectories are shown in the co-rotating frame, assuming an outflow of 600 km s^{-1} and an outward pulsar-wind acceleration of 12 times the pulsar gravitational attraction (required to give a reasonable scale). Orbital motion is clockwise. Such a system can give a sharp entry eclipse and a sharp refraction exit eclipse, but the exit eclipse would be geometrically unrelated in phase with the entry eclipse, and for refraction to be important the exit eclipse would also have a very large associated time delay. Eclipses are instead observed to be quite symmetric, with only a small time delay.

companion (the impact parameter at eclipse) differs from unity by less than one part in 10^4 , which is much too small to give a refractive eclipse. A refractive eclipse may result (without much time delay) only for a very steep gradient in electron concentration. This could indeed be provided by an advancing cometary structure (corresponding to the entry eclipse in Fig. 1), but would not necessarily give exit from the eclipse at precisely the symmetric phase angle, if eclipse was to come from refraction in the tail; furthermore, one would then expect much larger time delays than are observed.

On the other hand, such eclipses are perfectly consistent with occultation by plasma trapped in an extended magnetosphere. An extended magnetosphere is quite similar in morphology to a comet, with a tight 'nucleus' (the closed field lines) and an extended 'tail' (field lines dragged back by wind interactions). However, the trapped plasma concentration on the closed field lines can be, and usually is, orders of magnitude larger than that on the open field lines, where the plasma is flowing away and being lost. Thus a magnetosphere would present a symmetric occulting object (in contrast to an outflowing cometary flow), and the time delays would be attributable to the small concentrations in the tail (again in contrast to a cometary flow, for which the concentrations in the head cannot be dramatically different from those in the near tail). A significant mass loss does not arise obviously from such a model. The observed maximum column density of $5 \times 10^{16} \text{ cm}^{-2}$ is converted to a loss rate by multiplying by some characteristic transverse scale and some characteristic velocity. These are plausibly determined by, respectively, the eclipse size and the escape velocity ($\sim 3 \times 10^7 \text{ cm s}^{-1}$), which values then give a scale loss rate of 10^{12} g s^{-1} , which is four orders of magnitude smaller than is required for evaporation.

Given that the pulsar-wind field could be 10–30 G at the companion (if $\dot{P} \geq 10^{-20} \text{ s s}^{-1}$), one needs a comparable field

from the companion at the magnetopause, and hence a surface field of $\sim 10^5$ G, which value is not at all exceptional for white dwarfs. The magnetopause would be about 13 times the size of the companion, which is very nearly the case for the magnetosphere of the Earth and smaller than that of Jupiter.

If the time delay originates in magnetized plasma, there should be a differential time delay that depends on the sign of circular polarization. If the observing frequency (ω) is high compared with cyclotron (ω_c) and plasma (ω_p) frequencies, the time delay $\delta t_{\pm}$ is

$$\delta t_{\pm} = \frac{\omega_p^2}{2\omega^2} \frac{L}{c} \left[1 \pm \frac{2\omega_c}{\omega} \right]$$

(where L is the path length), so the difference between left- and right-circular-polarization time-delays is a factor of $4\omega_c/\omega$ of the average time delay ($\sim 400 \mu\text{s}$ at 430 MHz). The observed Faraday delay seems to be less than $\sim 2 \mu\text{s}$ (D. Stinebring, personal communication), implying that the limiting tail field is ~ 0.2 G.

This value is small compared with that at the magnetopause, whereas one expects that the two should be roughly comparable (within about an order of magnitude). If the magnetic field is inclined relative to the spin axis, however, it is possible that the line of sight is through both lobes of the tail (one with magnetic field away from the companion and the other towards the companion) so that the time delays average out. For the expected radial expansion, these time delays should be established by particles and fields within an orbital radius. Unfortunately the Faraday rotation would be too sensitive a measure of the time delays, giving large rotation measures that would effectively depolarize the radio signal, which is consistent with the observed weak, or possibly non-existent, polarization of the pulsar. Some dependence of frequency on eclipse size would also be expected, being determined by the electron concentration inside the magnetopause.

In the other limit of low magnetization we require a model with characteristics similar to those of Epsilon Aurigae. In ϵ Aur a cloud of scattering particles is thought to cause occultation, although there are difficult, unresolved questions concerning the particle dynamics, evolution, maintenance and origin. Such questions would also pertain here; however, we have determined numerically that it is at least possible for independent particle ('Kepler') retrograde orbits to exist that extend far enough out from the companion to cause the eclipse. Larger orbits quickly become unstable, suggesting that the size of an occulting cloud may be fixed by orbital stability. One problem common to ϵ Aur is the requirement for the disk-shaped cloud to be either inclined or thick, as there is a low *a priori* probability of occultation of the pulsar by a thin disk seen edge on. Too thick a disk, however, would obscure optical radiation from the companion. If indeed the companion had no significant magnetic field, the pulsar wind could flow directly onto its photosphere, lighting up the hemisphere facing the pulsar, as observed^{5,6}. At the terminator, plasma and entrained gas can be driven away from the companion. The natural consequence would be the formation of a meteor-like trail of ionized gas (which would presumably then produce the time delay) together with a 'fountain' of neutral gas with a wide velocity dispersion. If the companion is indeed the remnant of an evolved star, the vapour-phase fountain will contain some carbon (perhaps predominantly), which is readily condensed and highly refractory. Such a physical system could therefore support an orbiting condensate cloud, with solid particles accreting to significant sizes on each pass of their orbits through the gas fountain. Retrograde orbits are the natural expectation for low-velocity outflow, for which the tail in Fig. 1 is swept back so much that it encircles the companion. Particles several centimetres in size would be required to occult radio signals. About 10^{-10} of the companion's mass, in the form of such particles, would be

adequate to cause extinction. If such a disk reprocessed the energy intercepted by the companion, it would constitute a 20th magnitude infrared object at a temperature of $\sim 10^3$ K.

I acknowledge valuable discussions with D. Stinebring, C. R. O'Dell, D. D. Clayton, R. Smalley and A. J. Dessler. This research was supported by the NSF.

Received 19 August; accepted 28 November 1988.

1. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237-239 (1988).
2. Bailyn, C. *Nature* **334**, 298 (1988).
3. Phinney, E. S., Evans, C. R., Blandford, R. D. & Kulkarni, S. R. *Nature* **333**, 832-834 (1988).
4. Bowers, R. & Deeming, T. *Astrophysics J Stars* **260** (Jones and Bartlett, Boston, 1984).
5. Van Paradijs, J. *et al.* *Nature* **334**, 684-686 (1988).
6. Fruchter, A. S., Gunn, J. E., Lauer, T. R. & Dressler, A. *Nature* **334**, 686-689 (1988).

Evidence from the Semarkona ordinary chondrite for ^{26}Al heating of small planets

I. D. Hutcheon* & R. Hutchison†

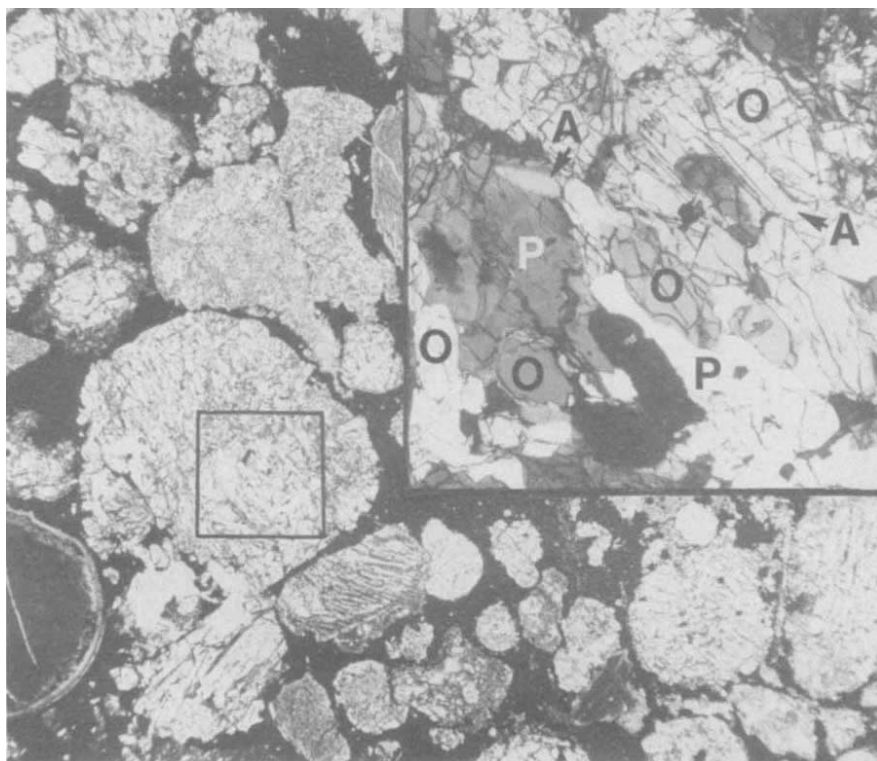
* The Lunatic Asylum of the Charles Arms Laboratory, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

† Mineralogy Department, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK

The radioactive decay of ^{26}Al , to form ^{26}Mg , may have constituted an important heat source for melting of small planetary bodies in the early Solar System. Although large excesses of radiogenic ^{26}Mg , arising from *in situ* decay of ^{26}Al (half life $\sim 7.2 \times 10^5$ yr), have been found in ten carbonaceous and ordinary chondritic meteorites^{1,2}, the distribution of ^{26}Al in the solar nebula and its importance as a heat source remain to be determined. All previous observations of excess ^{26}Mg are confined to Al-rich minerals in refractory, Ca,Al-rich inclusions (CAIs). An assessment of the distribution of ^{26}Al in the Solar System from these data is difficult because CAIs are only a minor constituent of chondritic meteorites associated with high-temperature nebular processes, and because the initial $^{26}\text{Al}/^{27}\text{Al}$ ratio inferred from the CAI data is not constant³ but varies from the commonly measured value of $\sim 5 \times 10^{-5}$ to values $< 1 \times 10^{-7}$. Here we report the first observation of radiogenic ^{26}Mg in non-refractory meteoritic material, a plagioclase-bearing, olivine-pyroxene clast chondrule in the Semarkona ordinary chondrite. The inferred initial abundance of ^{26}Al ($^{26}\text{Al}/^{27}\text{Al} = (7.7 \pm 2.1) \times 10^{-6}$) is sufficient to produce incipient melting in well insulated bodies of chondritic composition. We conclude that planetary accretion and differentiation must have begun on a timescale comparable to the half life of ^{26}Al and that, even if widespread melting did not occur, ^{26}Al heating played a significant role in thermal metamorphism on small planets.

If ^{26}Al was widespread in the nebula and was the heat source for melting small planets, some evidence of this should be preserved in 'old' igneous rocks that cooled on a timescale comparable to the mean life of ^{26}Al . No excesses of ^{26}Mg were found in earlier studies of plagioclase for equilibrated chondrites, achondrites and iron meteorites⁴, and from three unequilibrated ordinary chondrites². The preservation of isotope effects is strongly affected by the extent of thermal metamorphism, for which reason the present search for radiogenic ^{26}Mg has focused on the Semarkona chondrite, which is among the least metamorphosed chondrites⁵. Semarkona is a highly unequilibrated, LL-group ordinary chondrite which comprises chondrules and clasts set in a fine-grained matrix containing smectite, calcite and magnetite⁶. *In situ* hydrothermal alteration has selectively affected some chondrules and clasts, suggesting that Semarkona formed in an environment similar to that of the CM2 carbonaceous chondrites and is probably petrological type 2.

Fig. 1 Optical micrograph (plane-polarized light) of a polished thin section of the Semarkona chondrite. The clast chondrule, CC-1, is indicated by the square; note the roughly pentagonal outline. The square is 500 μm across. Inset shows an enlarged view of CC-1 contained within the square, as seen with crossed polars. Olivine dendrites (O) are poikilitically enclosed in pyroxene (P); anorthite (A) occurs as an interstitial phase.



Anorthite was found in a ~ 1 -mm diameter clast chondrule with a roughly pentagonal outline (Fig. 1). The chondrule, 1805-5 CC-1, is composed of olivine dendrites, some of which traverse the whole chondrule, intergrown with pigeonite. Pigeonite crystals range up to 200 μm in length and poikilitically enclose ovoid portions of olivine dendrites. Crystals of both minerals are truncated at the chondrule margin. The pigeonite may be zoned through subcalcic augite to augite. Anorthite is present as an interstitial phase, typically as $40 \times 100 \mu\text{m}$ laths. CC-1 is free from hydrothermal alteration and contains no glass; ilmenite and kamacite occur as trace constituents. Olivine has a minimum compositional range of Fo_{76} to $\text{Fo}_{72.5}$ (here Fo denotes forsterite); CaO and Cr_2O_3 contents are 0.28–0.32 and 0.28–0.55 wt%, respectively. Pyroxene has a minimum range of $\text{Wo}_{64}\text{En}_{36}$ to $\text{Wo}_{37}\text{En}_{48}$, with the corresponding range of 0.49 to 1.43 wt% Cr_2O_3 (here Wo and En denote wollastonite and enstatite respectively). Anorthite contains no detectable Na_2O or K_2O (less than 0.08 and 0.03 wt%, respectively) but contains 0.6–0.8 wt% FeO and 0.5–0.65 wt% MgO. Representative chemical analyses are given in Table 1.

The Mg isotope data were collected with the PANURGE IMS-3F ion probe in an automated peak-jumping mode at a mass-resolving power of 3,300, following techniques discussed by Hutcheon *et al.*⁷ Mass-dependent isotope fractionation (F_{Mg}) is calculated from the deviations in the $^{25}\text{Mg}/^{24}\text{Mg}$ ratios measured in olivine and pyroxene in CC-1 relative to ratios measured in terrestrial standards. Variations in the $^{26}\text{Mg}/^{24}\text{Mg}$ ratio are expressed as $\delta^{26}\text{Mg}$, relative to $^{26}\text{Mg}/^{24}\text{Mg} = 0.13955$, and are calculated after normalizing to $^{25}\text{Mg}/^{24}\text{Mg} = 0.12663$ to correct for instrumental mass fractionation. The $^{27}\text{Al}/^{24}\text{Mg}$ ratios were calculated from the $^{44}\text{Ca}^+/^{24}\text{Mg}^+$ ion ratios using a working curve calibrated against electron-probe measurements of terrestrial plagioclase. Abundances of the rare-earth elements (REE) in CC-1 were also measured with PANURGE using an energy-filtering technique⁸ at a resolving power of ~ 500 . Concentrations in olivine, pyroxene and anorthite were calculated from REE secondary-ion intensities normalized to the $^{42}\text{Ca}^+$ intensity. Pyroxene from the Angra dos Reis achondrite and a synthetic melilite glass⁹ were used as standards.

The Mg isotope data (Table 2 and Fig. 2) for CC-1 show small but clearly resolved excesses of ^{26}Mg in five anorthite crystals and isotopically normal Mg in olivine and pigeonite. There is no evidence for isotopically fractionated Mg in any phase, $|F_{\text{Mg}}| \leq 3\%$ AMU^{-1} . Excesses of ^{26}Mg (denoted $^{26}\text{Mg}^*$) in anorthite range from $\delta^{26}\text{Mg} = 2.5 \pm 1.6\%$ to $4.2 \pm 1.8\%$ and are linearly correlated with the $^{27}\text{Al}/^{24}\text{Mg}$ ratio measured at each point of analysis. The anorthite crystals exhibit only a limited range of Al/Mg ratios; an extensive search for anorthite with $\text{Al/Mg} > 80$ was unsuccessful. These data are characteristic of the *in situ* decay of ^{26}Al in a system with a uniform initial isotope composition, undisturbed by metamorphic processes. In particular, the line defined by the anorthite data is consistent with isotopically normal Mg, as found in olivine and pyroxene, for $^{27}\text{Al}/^{24}\text{Mg} \ll 1$. The data from the three coexisting phases define an excellent linear correlation on the Al–Mg evolution

Table 1 Chemical analyses of olivine, pyroxene, and plagioclase and the bulk composition of Semarkona CC-1

	1	2	3	4	5	6
SiO_2	37.4	37.4	51.2	48.7	44.9	46.0
TiO_2	—	—	0.40	0.54	—	0.2
Al_2O_3	—	—	2.35	4.35	34.3	4.2
Cr_2O_3	0.51	0.41	0.49	1.43	—	0.5
FeO	21.5	24.8	17.0	8.70	0.85	13.9
MnO	0.45	0.38	0.53	0.34	—	0.4
MgO	39.1	37.2	22.6	16.9	0.81	30.2
CaO	0.30	0.32	3.91	18.0	20.3	3.6
Sum	99.3	100.5	98.5	99.0	101.2	99.0
Fo	76.1	72.5	—	—	—	—
Wo	—	—	8.0	37.1	—	—
En	—	—	64.0	48.4	—	—
Fs	—	—	28.0	14.5	—	—

1, 2, extremes of olivine composition; 3, 4, extremes of pyroxene composition; 5, typical plagioclase; 6, bulk composition. Data are wavelength-dispersive electron-microprobe analyses except column 6, which is the average of 30 energy-dispersive analyses. Fs is ferrosilite.

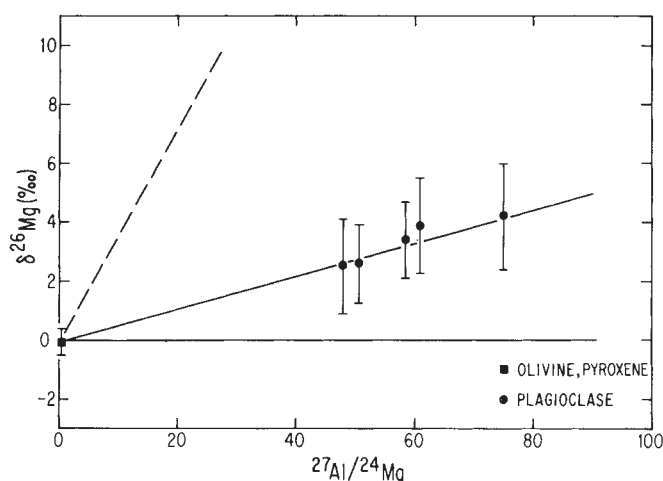


Fig. 2 Al-Mg evolution diagram for coexisting olivine, pyroxene and anorthite in the Semarkona chondrule CC-1. Anorthite contains small excesses of radiogenic $^{26}\text{Mg}^*$ clearly resolved from normal Mg. The linear correlation between $\delta^{26}\text{Mg}$ and Al/Mg is produced by the *in situ* decay of ^{26}Al ; the initial ^{26}Al abundance is given by the slope of the solid line through experimental points, $(^{26}\text{Al}/^{27}\text{Al})_0 = (7.7 \pm 2.1) \times 10^{-6}$. The dashed line with slope corresponding to $(^{26}\text{Al}/^{27}\text{Al})_0 = 5 \times 10^{-5}$ represents the *in situ* decay of ^{26}Al in typical Allende Type B CAI. The horizontal solid line represents normal Mg.

diagram (Fig. 2) with a slope corresponding to $^{26}\text{Mg}^*/^{27}\text{Al} = (7.7 \pm 2.1) \times 10^{-6}$.

The isotope data from CC-1 demonstrate for the first time that short-lived ^{26}Al was not confined to Ca,Al-rich refractory materials associated with early nebular condensation. This evidence for the *in situ* decay of ^{26}Al in a mafic chondrule from an ordinary chondrite raises two important issues: (1) Does the abundance of $^{26}\text{Mg}^*$ in CC-1 provide chronological information on the formation of chondrules? (2) Is the inferred abundance of ^{26}Al sufficient to cause melting of chondritic planetesimals? The initial $^{26}\text{Al}/^{27}\text{Al}$ ratio indicated by the CC-1 data is approximately a factor of six less than the value of $\sim 5 \times 10^{-5}$ which is characteristic of many Allende type B1 CAIs. If this difference is due to the decay of ^{26}Al from a uniform initial value, then Semarkona CC-1 formed $\sim 2 \times 10^6$ yr after many type B1 CAIs. However, the large variation in $^{26}\text{Mg}^*/^{27}\text{Al}$ ratios measured in CAIs³ strongly suggests that the $^{26}\text{Al}/^{27}\text{Al}$ ratio was heterogeneous in the solar nebula. Differences in inferred $^{26}\text{Al}/^{27}\text{Al}$ ratios cannot, therefore, be attributed to differences in age. The presence of 'live' ^{26}Al in CC-1 does require that the parent rock solidified very early, within 10^7 years of the time of the ^{26}Al production and not more than a few million years after refractory inclusions.

The interpretation of this evidence of the *in situ* decay of ^{26}Al in terms of the abundance of ^{26}Al in the solar nebula depends critically on the identification of CC-1 as a non-refractory chondrule or fragment produced by igneous activity on a planetary body. The modal mineralogy (olivine 59%, pyroxene 34%, anorthite 7%) and bulk composition (column 6, Table 1) clearly show that CC-1 does not consist of Ca,Al-rich refractory material. The angular form, with crystals terminating abruptly against (rather than nucleating on) the chondrule margin, the mineral chemistry (Table 1) and the exceptionally low Na/Al ratio indicate that CC-1 is a noritic clast chondrule of the poikilitic variety¹⁰, that is, an abraded fragment of an igneous rock and not a quenched liquid droplet. As a group, noritic chondrules have affinities to lunar ANT rocks and to the eucrites^{11,12} and most plausibly formed from volatile-poor, differentiated planetary material.

The abundance and distribution of rare-earth elements in CC-1 strongly support a planetary, rather than nebular, origin.

The bulk REE content of CC-1 is dominated by the contribution from pyroxene. The REE pattern in high-Ca pyroxene is strongly fractionated, with the heavy REEs (HREEs) enriched; chondrite-normalized abundances (except Eu) increase smoothly from ~ 0.4 for La to ~ 5 for Tm. Anorthite exhibits a complementary, HREE-depleted pattern with chondrite-normalized abundances (except Eu) less than unity and a large, positive Eu anomaly ($\text{Eu}/\text{Sm} \approx 38$). REE abundances in olivine are very low, ≤ 0.01 times chondritic levels for all REEs except La and Yb. The respective REE patterns are characteristic of the partitioning of REEs between basaltic liquid and crystallizing anorthite and pyroxene. The anorthite REE pattern is qualitatively similar to that of plagioclase from the Serra de Magé eucrite¹³, although the magnitude of the Eu anomaly in CC-1 anorthite is $\sim 30\%$ greater, comparable to Eu anomalies found in lunar anorthosites^{14,15}. The REE abundances in high-Ca pyroxene are in good agreement with predicted REE concentrations based on augite/liquid distribution coefficients^{16,17} for a parent liquid with levels of REEs about 8 times those in chondrites.

The mineralogical and chemical evidence argues convincingly for a planetary, rather than nebular, origin for CC-1. The presence of excess $^{26}\text{Mg}^*$ in CC-1 indicates that small planetary bodies accreted early and incorporated a substantial amount of ^{26}Al . If CC-1 therefore represents a fragment of early crystallizing material, melting and the production of differentiated liquids must have occurred on a timescale comparable to the mean life of ^{26}Al . During and after crystallization, cooling must have been rapid, and sub-solidus equilibration minimal; otherwise, olivines would have lost CaO and Cr_2O_3 (incorporated into olivine as the primary liquidus phase before pyroxene nucleation) to their neighbouring pyroxenes.

The magnitude of planetary heating caused by ^{26}Al depends primarily on the abundance of ^{26}Al and on the characteristic thermal diffusivity of the body, $\eta = \pi \kappa t / \rho C_p a^2$, where κ is the thermal conductivity, ρ is the density, C_p is the specific heat, a is the planetary radius and t is the time after accretion. For $\eta \ll 1$, the centre of the planet is effectively insulated and most of the heat is retained. In this limiting case, corresponding to $a \geq 30$ km for a compact, chondritic body, the maximum central temperature is linearly related to the ^{26}Al content⁴. For the ratio $^{26}\text{Al}/^{27}\text{Al} \approx 8 \times 10^{-6}$ inferred for CC-1, ^{26}Al heating will produce central temperatures of $\sim 1,100^\circ\text{C}$ in a chondritic body. This ratio refers to the ^{26}Al abundance at the time CC-1 became closed to Mg and Al diffusion, and represents a lower limit to the ^{26}Al abundance at the time at which planetary bodies formed. ^{26}Al will be much more abundant and central temperatures will be much higher in bodies that accreted earlier. Incipient melting will occur ($T \geq 1,200^\circ\text{C}$) in chondritic bodies accreting only $\sim 70,000$ yr earlier, and melting will be widespread in bodies formed more than 100,000 yr earlier.

Detailed solutions to the heat diffusion equation¹⁸ show that,

Table 2 PANURGE ion-probe analyses of the Mg isotope composition of olivine, pyroxene and plagioclase in Semarkona CC-1

Phase	$\Delta^{25}\text{Mg}$ (‰)†	$\delta^{26}\text{Mg}$ (‰)‡	$^{27}\text{Al}/^{24}\text{Mg}$
Plag 1	0.9 ± 1.2	2.5 ± 1.6	48 ± 2
Plag 3	2.5 ± 1.2	2.6 ± 1.4	51 ± 2
Plag 4	1.5 ± 1.3	3.5 ± 1.6	58 ± 2
Plag 5	1.8 ± 1.3	4.2 ± 1.8	75 ± 4
Plag 6	-2.5 ± 1.3	3.9 ± 1.7	61 ± 3
Olivine*	-6.1 ± 0.6	-0.3 ± 1.0	<0.001
Pyroxene*	-1.8 ± 0.6	0.3 ± 1.0	~ 0.15

* Averages of two analyses of two crystals each. All errors are 2σ .

† Deviation (‰) of unnormalized data relative to $^{25}\text{Mg}/^{24}\text{Mg} = 0.12663$.

‡ Deviations relative to $^{26}\text{Mg}/^{24}\text{Mg} = 0.13955$, data normalized to $^{25}\text{Mg}/^{24}\text{Mg} = 0.12663$.

even if ^{26}Al did not produce widespread melting, it was probably a dominant factor driving thermal metamorphism on small planets. For $^{26}\text{Al}/^{27}\text{Al} = 8 \times 10^{-6}$, the central 25-km-diameter region of a 40-km-diameter, solid chondritic body ($\kappa \approx 0.007 \text{ cm}^2 \text{ s}^{-1}$) will be heated to temperatures exceeding 500°C . Temperatures will remain at over $\sim 500^\circ\text{C}$ in the central region for $\sim 1.5\text{--}3.6 \times 10^6$ yr, depending on the distance from the centre. In a porous, loosely compacted body, the thermal diffusivity is much lower ($\kappa \leq 0.0002 \text{ cm}^2 \text{ s}^{-1}$) and temperatures exceeding 500°C are obtained in the central 8-km region of a 10-km-diameter body. Similarly, the low thermal diffusivity of a thick ($\sim 50\text{--}100\text{-m}$) regolith will reduce conductive heat loss and increase interior temperatures relative to the value given above for a solid body.

Estimates of metamorphic temperatures of ordinary chondrites range from $\sim 250\text{--}400^\circ\text{C}$ for type 3 to $\sim 750\text{--}950^\circ\text{C}$ for type 6. Sufficient heat should be generated by decay of ^{26}Al in the interiors of 10–100-km-diameter bodies that accreted early and incorporated ^{26}Al at the level inferred for Semarkona CC-1 to account for the majority of the metamorphic features in types 3, 4 and 5 ordinary chondrites.

Semarkona CC-1 is the first anorthite-bearing chondrule from an ordinary chondrite that we have examined for radiogenic $^{26}\text{Mg}^*$. If ^{26}Al was an important heat source, evidence of $^{26}\text{Mg}^*$ should be found in similar plagioclase-bearing chondrules in other unequilibrated ordinary chondrites. Hinton and Bischoff² failed to detect $^{26}\text{Mg}^*$ in plagioclase in Al-rich chondrules from three ordinary chondrites, but excess $^{26}\text{Mg}^*$ at the level present in Semarkona CC-1 would not have been detected in their study.

A more extensive survey of the Mg isotope composition of plagioclase-bearing chondrules and fragments in unequilibrated ordinary chondrites is required to establish the average initial abundance of ^{26}Al .

We thank D. A. Papanastassiou and G. J. Wasserburg for helpful comments, A. Kennedy for assistance with REE measurements, J. S. Wen for assistance with the heat-conduction calculations and the US National Museum for the loan of the Semarkona section. Thoughtful reviews were provided by R. Hinton and E. Zinner. Research supported by NASA.

Received 22 August; accepted 21 November 1988.

1. Hutcheon, I. D., Armstrong, J. T. & Wasserburg, G. J. *Lunar planet. Sci.* **XVII**, 372–373 (1986).
2. Hinton, R. W. & Bischoff, A. *Nature* **308**, 169–172 (1984).
3. Wasserburg, G. J. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M.) 703–737 (Univ. of Arizona Press, Tucson, 1985).
4. Schramm, D. M., Tera, F. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **10**, 44–59 (1970).
5. Sears, D. W. *et al. Nature* **287**, 791–795 (1980).
6. Hutchison, R. *et al. Geochim. cosmochim. Acta* **51**, 1875–1882 (1987).
7. Hutcheon, I. D., Armstrong, J. T. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **51**, 3175–3192 (1987).
8. Zinner, E. & Crozaz, G. *Int. J. Mass Spectr. Ion Proc.* **69**, 17–39 (1986).
9. Beckett, J. R. *et al. Lunar planet. Sci.* **XIX**, 49–50 (1988).
10. Dodd, R. T. *Meteorites: A Petrologic-Chemical Synthesis* (Cambridge Univ. Press, 1981).
11. Hutchison, R. & Bevan, A. W. R. in *Chondrules and Their Origins* (ed. King, E.) 162–179 (Lunar planet. Inst., Houston, 1983).
12. Prinz, M. & Keil, K. *Phys. Chem. Earth* **10**, 215–237 (1977).
13. Ma, M.-S. & Schmitt, R. A. *Meteoritics* **14**, 81–89 (1979).
14. Philpotts, J. A. *et al. Proc. lunar Sci. Conf. 4th* 1427–1436 (1973).
15. Lindstrom, M. M. *et al. Proc. lunar Sci. Conf. 8th* 2137–2151 (1977).
16. Shih, C. *et al. Proc. lunar Sci. Conf. 6th* 1255–1285 (1975).
17. Schnetzler, C. C. & Philpotts, J. A. *Geochim. cosmochim. Acta* **34**, 333–340 (1970).
18. Carslaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd Edn (Oxford Univ. Press, New York, 1959).

Formation of chemically active chlorine compounds by reactions of atmospheric NaCl particles with gaseous N_2O_5 and ClONO_2

B. J. Finlayson-Pitts, M. J. Ezell & J. N. Pitts Jr

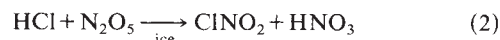
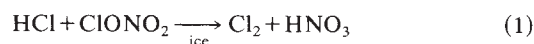
Department of Chemistry and Biochemistry,
California State University, Fullerton, California 92634, USA

Particulate NaCl and NaBr are known^{1,2} to react with NO_2 in air at concentrations of a few parts per 10^6 (p.p.m.) to form the corresponding nitrosyl halides XNO ($\text{X} = \text{Cl}, \text{Br}$), a process that may contribute to the photochemistry of polluted marine areas and remote regions such as the Arctic^{3,4}. Here we report experimental evidence that gaseous chlorine nitrate (ClONO_2) and dinitrogen pentoxide (N_2O_5) at p.p.m. concentrations also react with NaCl particles at 298 K, to form Cl_2 and ClNO_2 respectively. Thus, the chloride ion in a neutral salt can be transformed into gaseous, photochemically active forms of chlorine at 298 K in a manner analogous to that reported recently for the reactions of HCl in ice with ClONO_2 and N_2O_5 at 185–200 K (refs 5–8). This work suggests that reactions of N_2O_5 with NaCl in tropospheric marine areas, and possibly in cold climates where chloride salts are used on roads in the winter, should be considered in assessing the atmospheric chemistry of these regions. Furthermore, if the reactions of ClONO_2 and N_2O_5 with NaCl occur also at stratospheric temperatures and if NaCl is present in significant concentrations, these reactions may provide another mechanism for recycling ionic halide ions into gaseous species and hence should be considered in models of stratospheric ozone depletion.

Studies of the composition of particles collected in marine air environments, which are composed primarily of sea salt, frequently have a deficiency of chloride ions, as well as other halides such as Br^- , relative to sodium⁹. This has often been attributed^{10,11} to ion-exchange reactions of the atmospheric acids

HNO_3 and/or H_2SO_4 , forming HCl. More recently, NO_2 , at concentrations from a few thousand to a few p.p.m., was shown to react with NaCl and NaBr, forming the corresponding nitrosyl halides and sodium nitrate^{1,2,12,13}.

Recent investigations^{5–8} of heterogeneous reactions relevant to stratospheric ozone depletion, particularly in the Antarctic, have established that the following reactions, which are slow in the gas phase^{14–17}, occur rapidly when HCl is dissolved in ice:



These reactions serve to recycle the reservoir species ClONO_2 and HCl into more active forms of chlorine as a result of the rapid photolysis of the products Cl_2 and ClNO_2 in the Antarctic spring.

We show here that, in a manner analogous with reactions (1) and (2), ClONO_2 and N_2O_5 react readily with NaCl(s) at 298 K. This suggests that recycling of Cl^- from solids into photochemically active gaseous chlorine compounds is a general phenomenon, rather than being restricted to the HCl/ice system. This may have implications for the chemistry of both the troposphere and the stratosphere.

The reaction of NaCl with ClONO_2 , prepared using the method of Molina *et al.*¹⁸, was studied using both mass spectrometry (MS) and ultraviolet spectrometry (UVS). The MS apparatus is described in detail elsewhere^{1,2,9}. Mixtures of ClONO_2 (1.7–19%) in He (ultra-high purity, >99.995%) flowed through a needle valve into a low-pressure system operated at 0.11–0.7 torr. On the low-pressure side, the ClONO_2/He passed either through a glass bypass or through a cylindrical cell packed with NaCl. It then flowed into a mass spectrometer (dual-chamber Extrel EMBA II equipped with an 800-Hz tuning-fork chopper, lock-in amplifier and Teknivent Vector/One computer interface). Concentrations of ClONO_2 reacting with NaCl were $(3.8\text{--}11.9) \times 10^{14} \text{ cm}^{-3}$, corresponding to 15–48 p.p.m. if mixed with air at 1 atm. The residence time of the ClONO_2 in the NaCl-packed cell is estimated to be $\sim 5\text{--}10$ s.

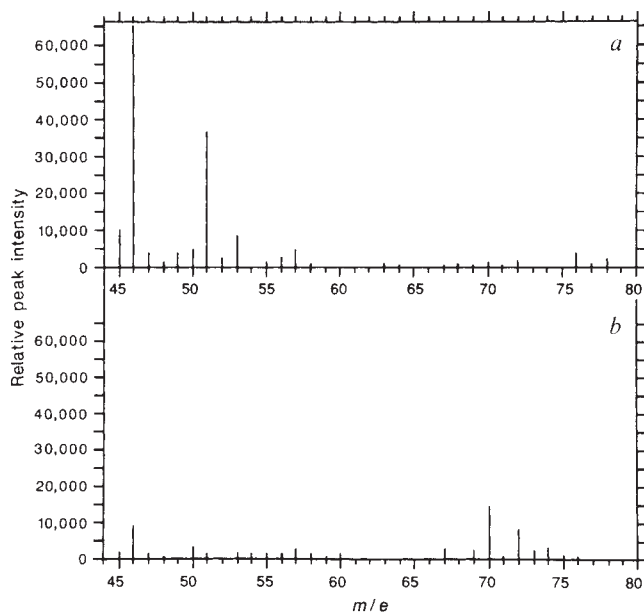


Fig. 1 Mass spectra of ClONO_2 when passed through the bypass (a), or passed over NaCl at 298 K (b). When ClONO_2 is exposed to NaCl , the peaks from ClONO_2 at $m/e = 46, 51$ and 53 drop significantly and those from Cl_2 at $m/e = 70, 72$ and 74 appear. Initial concentration of ClONO_2 was $7.5 \times 10^{14} \text{ cm}^{-3}$; total pressure of He was 0.112 torr. Intensities greater than 65,535 are off scale.

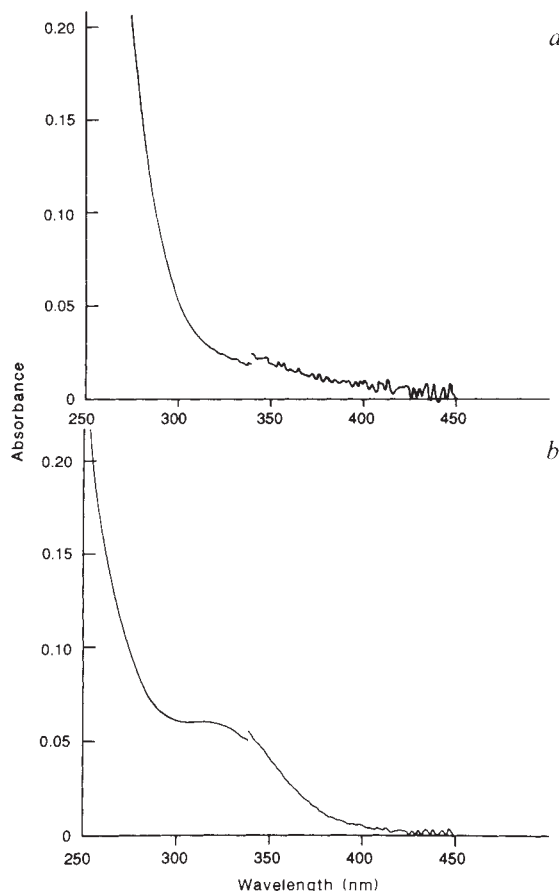


Fig. 2 Ultraviolet spectra of ClONO_2 : a, at 9.2 torr, unreacted, and b, at 3.8 torr ClONO_2 after reacting with NaCl at 298 K for 71 min. The increased absorbance at $\sim 320 \text{ nm}$ is due to the Cl_2 formed in the ClONO_2 - NaCl reaction.

In these and the N_2O_5 experiments reported below, the NaCl , which was present in excess compared with the gaseous reactants, was heated for several hours under vacuum before exposure to the gases; however, trace amounts of water may be present.

Figure 1 shows the mass spectra in the region $m/e = 44$ –80, obtained either when ClONO_2 bypasses the NaCl (Fig. 1a) or when it flows over the NaCl (Fig. 1b). The mass spectrum of the unreacted ClONO_2 in Fig. 1a shows peaks at $m/e = 46, 51$ and 53 resulting from fragmentation of the parent molecule; it is likely that NO_2 impurity also contributes to the $m/e = 46$ peak. No peaks are seen at $m/e = 70, 72$ or 74 , which are characteristic of Cl_2 ; in independent runs, the Cl_2 impurity in the ClONO_2 was estimated to be $<0.2\%$. When ClONO_2 passed over the NaCl , however, the peaks characteristic of ClONO_2 fell dramatically and those characteristic of Cl_2 appeared (Fig. 1b).

Further confirmation of the formation of Cl_2 was obtained using UVS. For these measurements 3.8 torr of ClONO_2 in 1 atm air (ultra-high purity, <0.01 p.p.m. total hydrocarbon content, <0.01 p.p.m. CO , <0.001 p.p.m. NO_x , <0.001 p.p.m. SO_2) was allowed to react with $\sim 100 \text{ g}$ NaCl in a 5-litre bulb for 71 min. The gas over the salt was then expanded into a 10-cm quartz cell and repressurized with air to 1 atm. Figure 2 compares the UVS spectrum (Fig. 2b) of the ClONO_2 after reaction with NaCl with that of a higher concentration of unreacted ClONO_2 (Fig. 2a). The increased absorbance at 320 nm after reaction is characteristic of Cl_2 . Using published absorption cross-sections¹⁹, it is estimated that $\sim 40\%$ of the ClONO_2 was converted to Cl_2 .

Thus both the MS and UVS studies imply the occurrence of the reaction



with enthalpy change $\Delta H_0^{298} = -82.8 \text{ kJ mol}^{-1}$.

In a second set of experiments, N_2O_5 was prepared by reacting NO_2 with excess O_3 and condensing the N_2O_5 at 195 K. Dilute mixtures of N_2O_5 in 1 atm air were expanded simultaneously into the reaction cell and into a 7-litre Fourier-transform infrared (FTIR) multi-pass cell with White cell optics (0.8 m base path, 40 m total pathlength), and both were repressurized to 1 atm with air. The spectrum of the initial N_2O_5 /air mixture was recorded using a Mattson Sirius 100 FTIR spectrometer at 2-cm^{-1} resolution with 64 scans. The reaction cell was a 5-litre bulb to which was attached a 200-ml bulb containing $\sim 100 \text{ g}$ of NaCl ; concentrations of N_2O_5 during the exposure were 2.9–216 p.p.m. The reaction bulb was rotated to thoroughly mix the salt with the gas and then allowed to react in the dark for ~ 1 hour. The reacted gas mixture was expanded into the FTIR cell, repressurized to 1 atm with air and the spectrum recorded.

Figure 3a gives a typical spectrum of the initial reactant mixture in the region $1,300$ – 700 cm^{-1} , showing the bands from N_2O_5 as well as those from some HNO_3 impurity. Figure 3b shows the spectrum of the gases after reaction with NaCl for 64 min. The reactant N_2O_5 has disappeared and bands from nitril chloride²⁰ (ClNO_2) are present, implying the reaction



for which $\Delta H_0^{298} = -55.2 \text{ kJ mol}^{-1}$. The yield of ClNO_2 was approximately proportional to the initial N_2O_5 concentration, as expected if N_2O_5 is the limiting reagent. When small amounts of impurity HNO_3 were present in the reactant, infrared bands from HCl centred around $2,885 \text{ cm}^{-1}$ were also observed, as expected for the HNO_3 - NaCl reaction^{10,11}. Hence, when acid is present, it reacts with NaCl in competition with N_2O_5 .

According to reactions (3) and (4), the second product of both reactions should be NaNO_3 . To confirm this, thin pressed plates were made from the NaCl which had been exposed to either N_2O_5 or ClONO_2 and the infrared transmission spectra recorded. Figure 4 shows the spectra of NaCl in the $1,500$ – 800 cm^{-1} region after exposure to N_2O_5 (Fig. 4a), after exposure

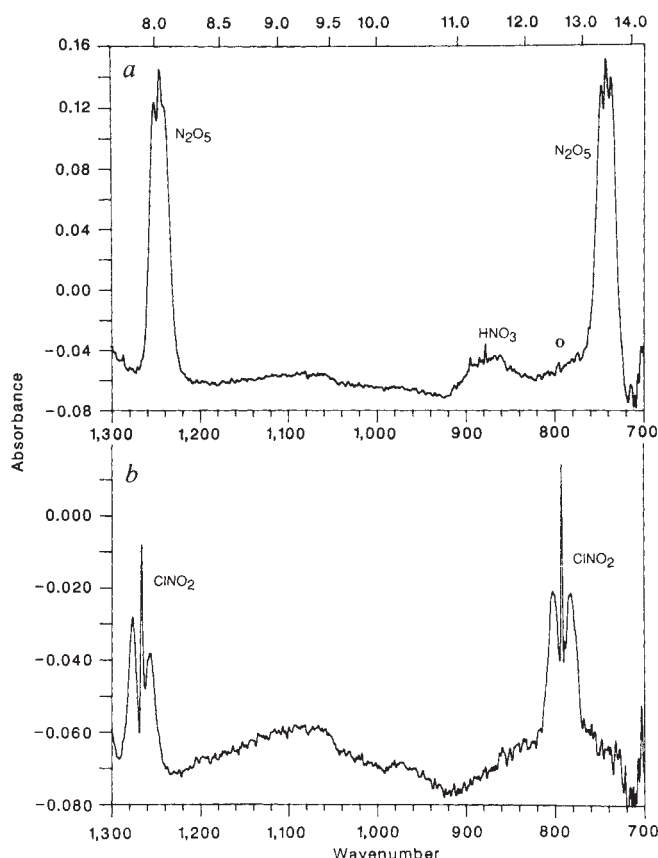


Fig. 3 Infrared spectra of 6.1 p.p.m. N_2O_5 in 1 atm air at 298 K, *a*, before and *b*, after reaction with NaCl for 64 min, showing that ClONO_2 is formed.

to ClONO_2 (Fig. 4*b*) and for comparison, the spectrum of a $\text{NaNO}_3/\text{NaCl}$ mixture (Fig. 4*c*). Clearly NaNO_3 is also a reaction product.

Thus, at room temperature both N_2O_5 and ClONO_2 react with solid NaCl in a manner analogous to reactions (1) and (2). In conjunction with our earlier work^{1,2}, which showed that NO_2 reacts with NaCl and NaBr to form ClNO and BrNO respectively, these studies indicate that recycling of halide ions in salts into gaseous, photochemically active forms can occur through a variety of reactions with species commonly encountered in the atmosphere. The fact that reactions (3) and (4) occur readily suggests that acid catalysis need not be invoked as a necessary condition for the conversion of inactive forms of chlorine into photochemically active gaseous species in either the troposphere or the stratosphere^{10,11,21,22}.

In the troposphere, the major source of NaCl in particles is as seasalt particles formed by wave action. Thus in polluted marine areas, where NO_2 , HNO_3 and N_2O_5 are present, production of ClNO , HCl and ClONO_2 may occur, in relative amounts that depend on the concentrations of the gas-phase species and the speeds of the individual reactions. Typical peak concentrations of NO_2 , HNO_3 and N_2O_5 in Los Angeles, for example, are ~ 100 parts per 10^9 (p.p.b.), 50 p.p.b. (ref. 9) and 15 p.p.b. (ref. 23; recent calculations (H. Biermann *et al.*, personal communication) suggest that this N_2O_5 value may be ~ 8 p.p.b.) respectively. The reaction efficiencies are not known, but we are currently measuring them. However, if the reaction of NaCl with N_2O_5 is approximately an order of magnitude faster than that with NO_2 , in heavily polluted marine environments it would compete with the NO_2 -NaCl reaction.

Both N_2O_5 and ClONO_2 hydrolyse on surfaces on which water is adsorbed^{9,24}; hydrolysis of N_2O_5 will thus compete with

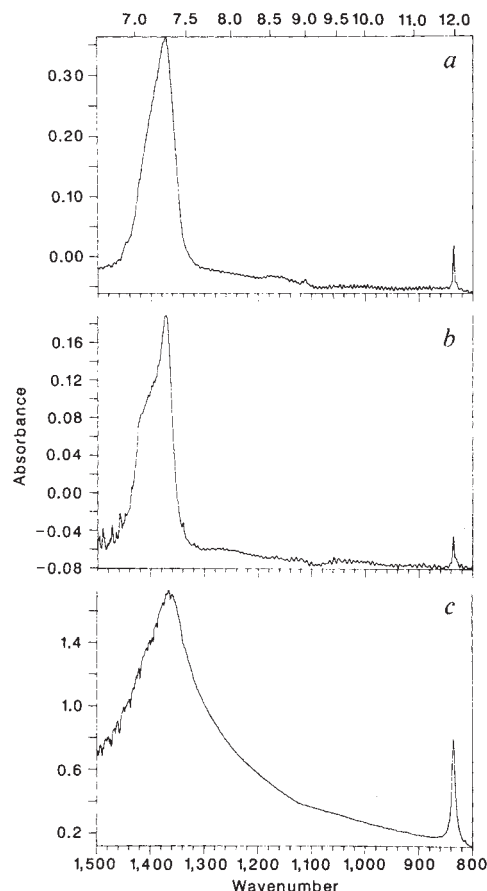
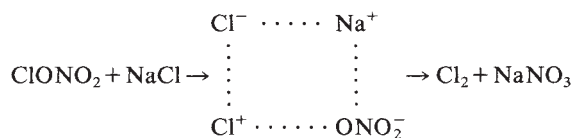


Fig. 4 Infrared spectra of NaNO_3 formed in the reactions of NaCl with 1.9 torr N_2O_5 (*a*) and NaCl with ClONO_2 (*b*) during mass spectrometry studies (see text). For comparison, *c* shows the spectrum of an authentic sample of a 12.6% $\text{NaNO}_3/\text{NaCl}$ mixture. All spectra referenced to unexposed NaCl.

its reaction with NaCl in the troposphere. Field measurements of NO_3 (which forms N_2O_5 and is in thermal equilibrium with it) indicate, however, that at relative humidities (RH) $\leq 50\%$, the lifetimes of NO_3 and hence N_2O_5 can be significant in the troposphere^{23,25}. The deliquescence point of NaCl is $\sim 75\%$, so that salt particles formed by wave action will be relatively dry at $\text{RH} \leq 50\%$ and hence under these conditions the N_2O_5 -NaCl reaction may be important.

All studies reported here were carried out at room temperature, but it seems reasonable that these reactions will continue to occur at a significant rate at lower temperatures. Chlorine nitrate is a polar molecule with the chlorine being the positive end; large dipole-dipole interactions with ionic NaCl should lead to the rapid formation of an ion-pair adduct which could rearrange as follows:



If there is no activation energy for the reaction, lower temperatures may increase the reaction efficiency by increasing the residence time of the ClONO_2 on the surface, as suggested by Molina *et al.*³ for the ClONO_2 -HCl/ice reaction.

If both reactions (3) and (4) do occur at significant rates at lower temperatures, they may play a role in the chemistry both of colder regions of the troposphere and of the stratosphere. For example, N_2O_5 could react with chloride salts used on roads

in winter, or indeed with frozen seasalt particles, to form ClONO_2 , which could then photolyse or undergo other secondary reactions. In addition, by analogy with the studies on $\text{NO}_2\text{-NaCl}$ and NaBr , other halide salts such as NaBr , if present, would be expected to undergo similar reactions; we are currently studying these systems. Such possibilities should be considered in models of tropospheric chemistry.

Whether NaCl is present in the stratosphere in significant quantities from such sources as meteor ablation or tropospheric injection is subject to some controversy²⁶⁻³⁰. If it is found, our work suggests that reactions (3) and (4) with ClONO_2 and N_2O_5 represent other possible mechanisms for recycling Cl^- into active forms. One key factor which may limit the importance of these reactions in recycling any stratospheric NaCl is that once the surface layer has been oxidized to NaNO_3 , this may effectively prevent further reaction. In contrast, in reactions (1) and (2) the HCl throughout the ice diffuses rapidly to the surface; thus all of the dissolved HCl , not only that on the surface, is available for reaction⁵⁻⁸.

There are several major caveats in extrapolating our studies to atmospheric, particularly stratospheric, conditions: they were carried out only at room temperature, in the absence of water or other co-pollutants and at higher concentrations than are found in the atmosphere. However, the mechanistic implications for the analogous reactions (1) and (2), which play key roles in stratospheric ozone loss particularly in the Antarctic, are sufficient to argue that the possible roles of reactions (3) and (4) in atmospheric chemistry should be investigated further. We are currently exploring the effects of temperature and water on reactions (3) and (4), and measuring the reaction efficiencies.

We are grateful to the NSF for support of this work, and to

Mr Stanley Johnson and Mr Frank Livingston for experimental assistance. We also thank Dr L. T. Molina for helpful discussions on the preparation of ClONO_2 .

Received 19 September; accepted 14 November 1988.

1. Finlayson-Pitts, B. J. *Nature* **306**, 676-677 (1983).
2. Finlayson-Pitts, B. J. & Johnson, S. N. *Atmos. Environ.* **22**, 1107-1112 (1988).
3. Sturges, W. T. *Atmos. Environ.* (in the press).
4. Barrie, L. A., Bottenheim, J. W., Schnell, R. C., Crutzen, P. J. & Rasmussen, R. A. *Nature* **334**, 138-141 (1988).
5. Molina, M. J., Tso, T.-L., Molina, L. T. & Wang, F. C.-Y. *Science* **238**, 1253-1257 (1987).
6. Tolbert, M. A., Rossi, M. J., Malhotra, R. & Golden, D. M. *Science* **238**, 1258-1260 (1987).
7. Leu, M.-T. *Geophys. Res. Lett.* **15**, 17-20 (1988).
8. Tolbert, M. A., Rossi, M. J. & Golden, D. M. *Science* **240**, 1018-1021 (1988).
9. Finlayson-Pitts, B. J. & Pitts, J. N. *Jr Atmosphere Chemistry: Fundamentals and Experimental Techniques* (Wiley, New York, 1986).
10. Robbins, R. C., Cadle, R. D. & Eckhardt, D. L. *J. Met.* **16**, 53-56 (1959).
11. Cadle, R. D. & Robbins, R. C. *Discuss. Faraday Soc.* **30**, 155-161 (1960).
12. Schroeder, W. H. & Urone, P. *Envir. Sci. Technol.* **8**, 756-758 (1974).
13. Chung, T. T., Dash, J. & O'Brien, R. J. *9th Int. Cong. Electron Microsc.* **440-441** (1978).
14. Molina, L. T., Molina, M. J., Stachnik, R. A. & Tom, R. D. *J. phys. Chem.* **89**, 3779-3781 (1985).
15. Hatakeyama, S. & Leu, M.-T. *Geophys. Res. Lett.* **13**, 1343-1346 (1986).
16. Friedl, R. R., Goble, J. H. & Sander, S. P. *Geophys. Res. Lett.* **13**, 1351-1354 (1986).
17. Atkinson, R. A., Aschmann, S. M., Tuazon, E. C., Goodman, M. A. & Winer, A. M. *J. atmos. Chem.* **5**, 83-90 (1987).
18. Molina, L. T., Spencer, J. E. & Molina, M. J. *Chem. Phys. Lett.* **45**, 158-162 (1977).
19. DeMore, W. B. *et al. Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling, Evaluation No. 8*, JPL Publ. no. 87-41, 15 September 1987.
20. Bernitt, D. L., Miller, R. H. & Hisatsune, I. C. *Spectrochim. Acta* **23A**, 237-248 (1967).
21. Crutzen, P. J. & Arnold, F. *Nature* **324**, 651-655 (1986).
22. Wofsy, S. C., Molina, M. J., Salawitch, R. J., Fox, L. E. & McElroy, M. B. *J. geophys. Res.* **93**, 2442-2450 (1988).
23. Atkinson, R. A., Winer, A. M. & Pitts, J. N. *Jr Atmos. Environ.* **20**, 331-339 (1986).
24. Rowland, F. S., Sato, H., Khwaja, H. & Elliott, S. M. *J. phys. Chem.* **90**, 1985-1988 (1986).
25. Platt, U., Winer, A. M., Biermann, H., Atkinson, R. & Pitts, J. N. *Jr Envir. Sci. Technol.* **18**, 365-369 (1984).
26. Rowland, F. S. & Rogers, P. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2737-2739 (1982).
27. Lamb, J. J. & Benson, S. W. *J. geophys. Res.* **91**, 8683-8689 (1986).
28. Ager, J. W. III & Howard, C. J. *Geophys. Res. Lett.* **13**, 1395-1398 (1986).
29. Shedlovsky, J. P. & Paisley, S. *Tellus* **18**, 499-503 (1966).
30. Delany, A. C., Shedlovsky, J. P. & Pollock, W. H. *J. geophys. Res.* **79**, 5646-5650 (1974).

Image processing using light-sensitive chemical waves

L. Kuhnert*, K. I. Agladze† & V. I. Krinsky†

* Max-Planck-Institut für Ernährungsphysiologie, Rheinland damm 201, D4600 Dortmund 1, FRG, formerly with Zentralinstitut für physikalische Chemie der Akademie der Wissenschaften der DDR

† Institute of Biophysics, USSR Academy of Science, 142292 Pushchino, Moscow Region, USSR

Image processing is usually concerned with the computer manipulation and analysis of pictures¹. Typical procedures in computer image-processing are concerned with improvement of degraded (low-contrast or noisy) pictures, restoration and reconstruction, segmenting of pictures into parts and pattern recognition of properties of the pre-processed pictures. To solve these problems, digitized pictures are processed by local operations in a sequential manner. Here we describe a special light-sensitive chemical system, a variant of the Belousov-Zhabotinskii medium, in which chemical reaction fronts ('chemical waves') can be modified by light. Projection of a half-tone image on such a medium initiates a very complex response. We are able to demonstrate contrast modification (contrast enhancement or contrast decrease up to contrast reversal from positive to negative and vice versa), discerning of contours (for example, segmenting of pictures up to the extreme case of skeletonizing) and smoothing of partially degraded pictures.

This light-sensitive chemical medium clearly works as a special image processor, and in contrast to sequential computer processes it can be considered to operate as a parallel process. The chemical medium used is a light-sensitive variant of the well-known Belousov-Zhabotinskii (BZ) reaction²⁻⁴. This reaction is the best studied example of self-organization in non-linear chemical systems operating far from equilibrium. In homogeneous well stirred solutions, temporal oscillations, bista-

Table 1 Compositions of the light-sensitive BZ reagent

	Composition 1 (mol l ⁻¹)	Composition 2 (mol l ⁻¹)
NaBrO_3	0.3	0.3
H_2SO_4	0.6	0.4
Malonic acid	0.2	0.2
Bromomalonic acid	0.1	0.1
$\text{Ru}(\text{bipy})_3\text{Cl}_2$	0.0025	0.0025

bility and deterministic chaos are observed. In distributed systems, however, if the reaction proceeds in a thin layer of reagent, spontaneous breaking of spatial symmetry occurs. Spatially propagating chemical-reaction fronts (so-called chemical waves) are observed, and two types of waves have been distinguished. If the reagent is in an oscillatory regime and the period of oscillations depends on the spatial coordinate, then a 'phase wave' spreads over the layer. Its velocity is inversely proportional to the phase gradient, as is the case in other physical phase-wave phenomena. The other type of wave, the 'trigger wave', is generated through autocatalytic feedback related to diffusion in the excitable regime. Trigger waves propagate with constant velocity according to a reaction-diffusion mechanism. Chemical trigger waves can be considered as analogous to propagation of flame fronts or of a grass fire. Trigger waves may also occur in the oscillatory regime if a very high phase gradient has been established, as in the case of a chemical system with a high concentration gradient of chemical substances. In such a case the velocity of phase waves is very low and trigger waves can dominate.

The BZ reaction occurs between bromate ions and an organic substrate (preferably malonic acid) in acidic conditions, in the presence of a metal-ion redox catalyst⁵. In the light-sensitive variant, the ruthenium bipyridyl complex $\text{Ru}(\text{bipy})_3^{2+}$ is used as the catalyst. The main steps in the complicated reaction network of the BZ reaction are autocatalytic oxidation of the catalyst and reduction of the oxidized catalyst by the brominated organic



Fig. 1 Light-sensitive BZ reaction (composition 1; Table 1) with a negative image projected onto it. Frame 1: appearance of an image; 2, 3: positive image; 4, 5: negative image; 6: transient phase. The cycle shown here then recurs.



Fig. 2 Image-contour discernment. Same projection as in Fig. 1, but using reagent composition 2 (Table 1). Frames 1, 2: positive image; 3, 4: the dark portion of the image acquires contours; 5-8: erosion of the contour images by trigger waves, which is the mechanism of skeletonizing; 9, 10: formation of a new contour outside the former retracting one; 11, 12: the two retracting contours now fit into one another.

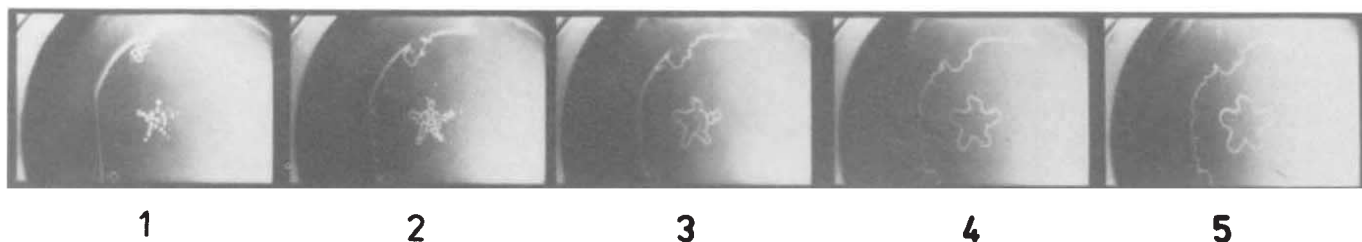


Fig. 3 Smoothing of a discontinuous image. The image, consisting of separate black points against a bright background, was projected on a Petri dish containing the light-sensitive BZ component of composition 2. Frame 1: an image appears as light points against the dark background; 2: each point becomes a source of a propagating concentric wave; 3: waves merge, producing a smoothed image; 4, 5: the image grows in size, and the degree of smoothing increases.

substrate. In the reduction step, liberated bromide ions act as inhibitors to autocatalytic oxidation; thus, bromide is the most important control intermediate and its concentration controls the reaction regime^{5,6}. The light-sensitive Ru-catalysed system offers the possibility of selecting the rate of bromide production by photochemical means. By illuminating a distributed system we can control the propagating effects, and, for particular light intensities, observe bifurcations between several reaction regimes. As a consequence, starting in an oscillatory regime and illuminating with increasing light intensity, we can control the velocity of phase waves, transform phase waves into trigger waves, control the velocity of trigger waves and, at very high light intensity, inhibit propagating phenomena completely. By projecting a half-tone image onto a thin layer of reagent, the light intensity is varied as a function of optical density and spatial coordinate. The chemical-wave phenomena are influenced accordingly, and several reaction regimes (for example, phase waves and trigger waves) may be observed in the same layer.

Figures 1-3 show some examples of time sequences of image processing performed by such a system. Fig. 1 demonstrates the possibility of controlling the contrast of an image, Fig. 2 illustrates contour delineation and Fig. 3 shows smoothing of discontinuous images. In these experiments a negative image was projected onto a thin layer (0.8 mm) of the reagent contained in a Petri dish of 9 cm diameter maintained by thermostat at 12 °C. Two reagent compositions were used, distinguished only by acidity (Table 1); composition 1 was used for contrast control and composition 2 for contour discernment and image smoothing. The time interval between the frames in Figs 1-3 was 10 s. In the initially homogeneous solution, there appeared first an inhomogeneity resulting from the influence of the projected image (frame 1 in Figs 1 and 2) and then the positive image itself (frame 2, Figs 1 and 2). For composition 1 (the more acidic solution, in which the oscillatory regime is more pronounced), projection of the negative causes only a change in the oscillatory period, whereas for composition 2 (lower acidity) the same projection causes a transformation from the oscillatory regime

(phase waves) to the excitable regime (trigger waves) in the brightest parts of the image, with the result that contours may be discerned.

In Fig. 1, the change in contrast of the image is accompanied by periodic positive-to-negative transitions, and in Fig. 2 the image contour is revealed. In the latter case, the contour of the brightest part of the image appears first (frame 3), and is then seen to shrink towards the centre of the image (frames 4-8). The process recurs periodically, creating a set of contours inscribed in one another and shrinking successively.

Projecting an image containing discontinuities initiates waves that merge during propagation to cause smoothing of the image (Fig. 3). During the exposure time (t), discontinuities with a length less than $l = \frac{1}{2} vt$ (where v is the propagation velocity of concentration waves in the chemical reaction) will become smoothed.

Let us consider the simplest mechanism of image processing by this light-sensitive reaction. Like the BZ reaction itself, this reaction has an oscillatory character; each point of the system alternates between a fluorescent blueish and an orange colour with a period of $\sim 1-2$ min. If a particular arrangement of blue and orange regions is prearranged in the reaction layer (by drawing on the reagent surface with a silver wire), then positive-negative image transitions will be observed (the orange colour being replaced by blue and vice versa) with a period equal to the reaction auto-oscillation period. Such an orange-blue configuration can alternatively be prearranged by projecting onto the medium a half-tone negative (Fig. 4a, top). Initially, no image appears in the reaction layer when projection commences, despite the spatial differences in illumination intensity (phase t_1 in Fig. 4). As time progresses, then because the reaction is light-inhibited (and therefore the auto-oscillation period is greater in the illuminated areas than in the non-illuminated areas), the change in colour occurs first in the non-illuminated areas (phase t_2); the initial negative image changes to a positive one. Then, because of auto-oscillations, the reverse takes place (phase t_3), whereupon the process repeats periodically.

Such transitions can be seen in Fig. 1 for a real image with a continuous distribution of illumination intensity. Frame 1 corresponds to the termination of phase t_1 , frames 2 and 3 to phase t_2 , frames 4 and 5 to phase t_3 , and frame 6 again to phase t_1 . In this case the contrast changes smoothly from a maximum through zero to the negative region (negative image).

By increasing the range of illumination intensity (or by changing the reagent composition), it is possible to discern the image contours (Fig. 2). In the simplest case, a contour is discerned which is a boundary between auto-oscillating and excitable regions, as illustrated in Fig. 4b. The concentration of bromide, c , which inhibits autocatalytic oxidation, determines the reaction mode: for $[Br^-] > c_2$ (Fig. 4b), the propagation effects are completely inhibited; for $c_1 < [Br^-] < c_2$, the medium is excitable, and for $[Br^-] < c_1$, the colour oscillates with a period equal to the reaction period (oscillatory regime). In the intermediate regime ($c_1 < [Br^-] < c_2$), propagating waves appear at the boundary of the auto-oscillating and excitable regions (Fig. 4b, bottom). A more detailed theoretical description of bifurcations in this reaction system, within the framework of a modified version of the well-known Oregonator model, will be given elsewhere⁷.

Note that chemical waves propagating to the centre of an image (Fig. 2, frames 9-12) can be used to discern the image skeleton. Such waves can also be used for image smoothing. Fig. 3 illustrates different stages in the development of a wave pattern resulting from the projection of a fragmentary image. In this case the non-illuminated points serve as leading centres for the creation of new waves, which merge during propagation and result in smoothing of the image. In illuminated areas, trigger-wave propagation is possible but not nucleation of new leading centres; light inhibits the nucleation of new waves.

These results are remarkable for several reasons. First, this system models some operations used in computer image-

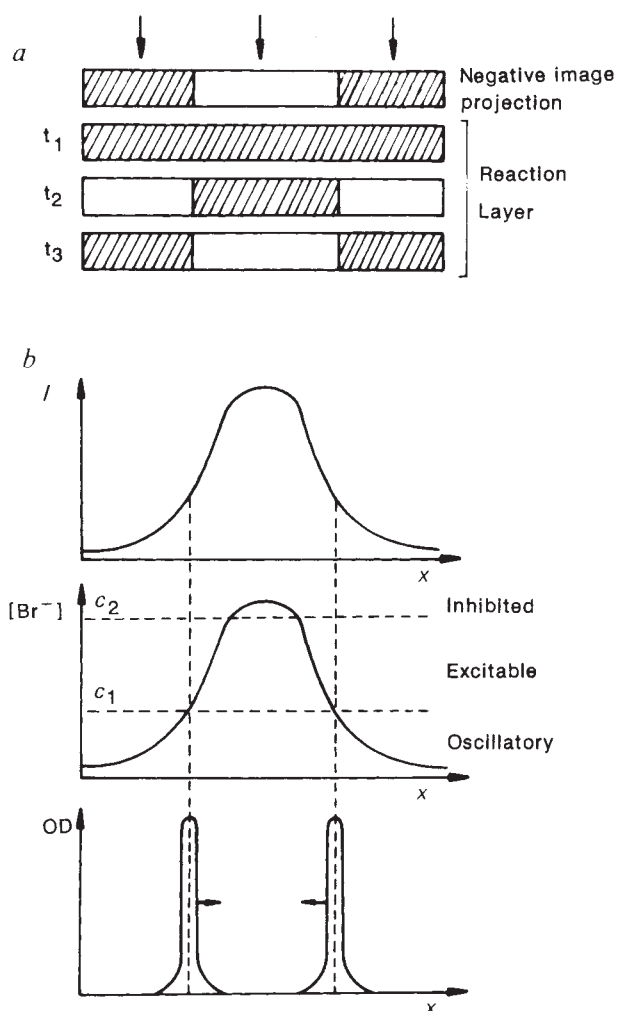


Fig. 4 Mechanism of image processing. *a*, Positive-negative transitions and contrast control. At the top is the negative image projection, and below, three different phases of the reaction (see text). *b*, Contour discernment. Top, illumination intensity of the projected image (I) and corresponding bromide concentration, as functions of the spatial coordinate. Bottom, optical density (OD) in the reaction layer, for the situation in Fig. 2, frame 4 (contouring), which is followed by skeletonizing (frames 5-8).

processing, and might suggest new algorithms (for example, chemical waves can be modelled by cellular automata). Second, there are some speculative connections to visual perception. Contrast enhancement in visual systems is based on the mechanism of lateral inhibition (Mach bands)⁸, for example, by neuronal interaction of neighbouring sensor cells. This kind of interaction can also be modelled by molecular diffusion within the framework of a particular chemical model with an autocatalytic step. Third, light-sensitive chemical waves can be seen as a chemical-hardware realization of certain ideas related to new computational systems such as synergetic computers or other network models^{9,10} (although in the system described here the interactions are only local). For example, the image-smoothing process (Fig. 3) is a realization of an associative memory device: from only an incomplete set of data, the whole pattern may be reconstructed⁹. To construct more sophisticated realizations we require more complex interactions, and in particular long-range interactions. These could be achieved by using special output devices that measure chemical concentration fluctuations at several selected points in the layer. In this way a learning network might be realized by chemical means.

Received 20 January; accepted 5 December 1988.

1. Rosenfeld, A. & Kak, A. C. *Digital Picture Processing* (Academic, New York, 1982).
2. Kuhnert, L. *Nature* **319**, 393–394 (1986).
3. Kuhnert, L. *Naturwiss.* **73**, 96–97 (1986).
4. Kuhnert, L., Pohlmann, L., Krug, H.-J. & Wessler, G. in *Selforganization by Nonlinear Irreversible Processes* (eds Ebeling, W. & Ulbricht, H.) (Springer, Berlin, 1986).
5. Zhabotinskii, A. M. *Concentrational Auto-oscillations* (Nauka, Moscow, 1974).
6. Field, R. J., Körös, E. & Noyes, R. M. *J. Am. chem. Soc.* **94**, 8649–8664 (1972).
7. Krug, H.-J., Pohlmann, L. & Kuhnert, L. *J. phys. Chem.* (submitted).
8. Ratcliff, F. *Mach bands* (Holden Day, San Francisco, 1965).
9. Haken, H. *J. chim. Phys.* **84**, 1289–1294 (1987).
10. *Computational Systems, Natural and Artificial* (ed. Haken, H.) (Springer, Heidelberg & New York, 1987).

Depletion of soil aluminium by acid deposition and implications for acid neutralization

J. Mulder, N. van Breemen & H. C. Eijck

Department of Soil Science and Geology, Agricultural University,
Box 37, 6700 AA Wageningen, The Netherlands

In acid soils acid rain is often largely neutralized by dissolution of aluminium^{1–3}, which is potentially phyto-toxic⁴. Here we show that only a minor fraction of soil aluminium is readily dissolved. This most soluble fraction consists largely of non-silicate, organically bound aluminium, which has been formed in the course of soil development. The current rapid and irreversible depletion of this fraction constitutes a drastic change in soil genesis. Depletion may eventually result in reduced acid neutralization, as a consequence of decreased dissolution of aluminium, and should be accounted for in soil acidification models.

Dissolution of aluminium in acid soils is one of the most pronounced soil chemical effects of acidic deposition^{1–3}. Aqueous aluminium is toxic to crop roots^{4–6} and has also been associated with forest die-back⁷. Because of its toxicity the soil chemical behaviour of aluminium has been a major topic in agronomic^{8,9} and ecological¹⁰ studies. Most of these studies deal with short-term phenomena such as speciation, exchange characteristics and solubility control. Long-term effects on aqueous aluminium in soils, resulting, for example, from changes in the soil solid phase, are generally ignored. In many

acid, sandy soils, however, an appreciable portion of the pools of various solid-phase aluminium fractions is currently leached from the soil. Table 1 indicates that leaching could lead to a significant depletion of certain soil aluminium fractions within several decades. The current depletion rates of extractable aluminium are significantly higher than the rates of aluminium transfer from silicate-bound to extractable soil aluminium. Maximum annual rates of formation of extractable aluminium amount to 0.030 mmol per kg of soil, as estimated from accumulation rates in the soil solid phase¹¹ (J.M. *et al.*, preprint) and from current rates of organic-aluminium precipitation¹¹. Decreasing content of various fractions of solid-phase aluminium could affect the aqueous aluminium concentrations. Insight into these aspects is essential for evaluating long-term effects of acid deposition.

Data on the sources of dissolved aluminium were obtained from a laboratory leaching experiment. Air-dried, <2 mm fractions of soil samples from the Netherlands and New Hampshire, USA, were taken from the surface and sub-surface horizons of two podzols (Haplorthods) and of a recent driftsand (Udipsamment) (Table 1). Duplicate samples of each soil horizon (5.0 g) were leached 10, 25 and 50 times, respectively, with 100 ml aqueous hydrochloric acid at pH 3.0. After each acid addition the suspensions were kept in tightly capped centrifuge tubes at 20 ± 2 °C for 24 hours. Suspensions were shaken manually three times in each 24 hours. The suspensions were centrifuged at 2,000 r.p.m. for 15 min, and 80 ml of the clear supernatants was decanted and replaced by 80 ml of fresh HCl solution. Sub-samples of each supernatant were pooled to give a single extract after 10, 25 and 50 HCl treatments. In each case, pooled leachates were analysed for all major solutes, dissolved aluminium concentrations were measured after 6, 24 and 72 hours in one additional leaching, and soil samples were freeze-dried and residual solid aluminium fractions were estimated by sequential selective dissolution (Table 1). Dissolved aluminium was assayed colorimetrically using pyrocatechol-violet.

We selected a leaching solution of pH 3.0, because this is close to the current pH of the water infiltrating many surface soils of forests and heathlands in north-western Europe³ (J.M. *et al.*, submitted). We enhanced aluminium mobilization rates relative to those in the field by applying HCl at a rate of 20 mmol H⁺ per kg of soil per 24 hours. This is considerably higher than the current acid load in the field (4 mmol H⁺ kg⁻¹ yr⁻¹).

Table 1 Pools of solid-phase aluminium in three acid, sandy soils from the Netherlands, and one from New Hampshire (USA), before laboratory leaching treatments, and annual aluminium mobilization rates from the respective soil layers in the field

Site	Soil	Depth horizon (cm)	Extractable aluminium*			Total aluminium†	Annual removal of aluminium‡
			KCl	Na ₄ P ₂ O ₇	(NH ₄) ₂ C ₂ O ₄ (mmol kg ⁻¹)		
Gerritsfles	driftsand	0–10 C	2.2	33.3	3.7	450	0.5
		10–40 C	0.7	35.7	6.3	427	0.3
Tongbersven	podzol	0–12 E	1.3	5.0	1.5	101	0.1
		12–35 B	13.7	196	3.3	446	0.4
Hasselsven§	podzol	0–7 E	11.1	21.2	7.7	221	0.5
		7–22 B	16.3	55.8	13.5	273	0.1
Hubbard Brook	podzol	0–5 E	1.2	7.8	3.0	1660	0.005
		15–30 B	5.8	322	51.1	2185	–0.004

* Solid-phase aluminium was extracted using selective dissolution. Soil samples (1 g) were sequentially extracted with 50 ml 1M KCl (0.5 hour), 50 ml 0.1M Na₄P₂O₇ (16 hours) and 50 ml 0.2M (NH₄)₂C₂O₄ adjusted to pH 3.0 (4 hours in the dark)¹¹. Extracted amounts represent exchangeable aluminium (KCl), organic aluminium and minor amounts of non-crystalline hydrous oxides (Na₄P₂O₇) and most amorphous aluminium hydroxides ((NH₄)₂C₂O₄)^{16,17}. Extracted aluminium was measured using colorimetry with pyrocatechol violet (KCl), inductive coupled plasma spectroscopy (Na₄P₂O₇), and atomic absorption spectroscopy with a nitrous oxide-acetylene flame ((NH₄)₂C₂O₄).

† Total aluminium, largely originating from aluminium silicates, was determined by X-ray fluorescence spectroscopy of a lithium tetraborate melt of the soil material.

‡ Calculated from bulk density (kg m⁻³) (H. F. van Dobben and J. Mulder, manuscript in preparation), and estimated aluminium fluxes (kmol ha⁻¹ yr⁻¹) (J.M. *et al.*, preprint). Monthly aluminium fluxes were estimated by multiplying the simulated soil water flux and the measured monthly dissolved-aluminium concentration in the soil solutions collected from suction plates in the field. Values for inorganic aluminium in the Hubbard Brook soils were taken from Driscoll *et al.*¹¹. Negative numbers indicate immobilization.

§ Not included in the leaching study.

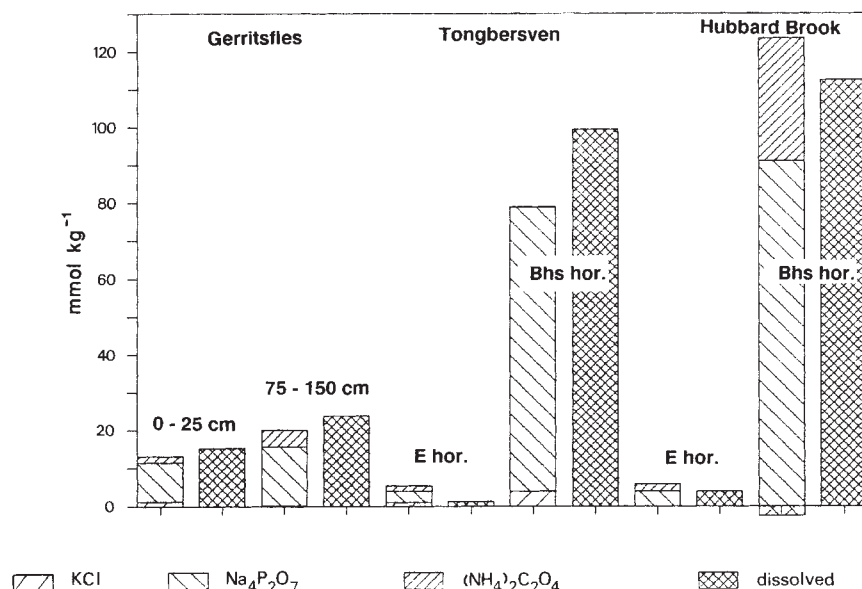


Fig. 1 The decrease in solid-phase aluminium, after 25 leachings with HCl (left bar, hatched) and the amount of aluminium removed in the leachates (right bar, cross-hatched) for six soil samples. Values are means of duplicates. Duplicates of dissolved aluminium were within 37% (E horizons) or 6% (all other horizons) of the mean; duplicates for the depletion of the $\text{Na}_4\text{P}_2\text{O}_7$ -extractable fraction were within 38% (E horizons) or 22% (other horizons) of the mean. Note that in the Hubbard Brook Bhs horizon KCl-extractable aluminium increases slightly over the course of 25 leaching cycles. This increase is due to the strong increase in dissolved aluminium relative to that in field conditions¹¹.

The cumulative amount of aluminium dissolved by HCl is invariably close to the amount of extractable ('free') aluminium removed (Fig. 1). Our data indicate that the dissolution rate of 'free' aluminium is high relative to that of the more abundant aluminosilicates (Table 1). Organically complexed ($\text{Na}_4\text{P}_2\text{O}_7$ -extractable) aluminium is the fraction that dissolves most, whereas the exchangeable fraction (KCl-extractable) contributes little to the total dissolved aluminium (Fig. 1). Thus, aluminium is transferred to solution mainly from the organically complexed solid phase, even when the pool of organic aluminium is small.

On addition of HCl, dissolved aluminium concentrations invariably rose to constant values within 24 hours, indicating attainment of chemical equilibrium. Charge balances of the leachates indicate that dissolved aluminium is present mainly as aquo- Al^{3+} . Only in the leachates of the podzol Bhs horizons is a significant fraction (20–30%) of dissolved aluminium organically complexed. (The podzol Bhs is a soil layer rich in organic-aluminium precipitates.) Dissolved aluminium concentrations are significantly correlated with the organic ($\text{Na}_4\text{P}_2\text{O}_7$ -extractable) aluminium content of the soil sample (Fig. 2). This further indicates the importance of solid aluminosilicates as a major source of dissolved aluminium. Because dissolved aluminium concentrations in our leachates increased rapidly to stable levels, but were highly undersaturated with respect to gibbsite ($p\text{Al}-3p\text{H}=-8.11$; ref. 13 and Fig. 3), exchange equilibrium between dissolved and adsorbed aluminium as described for forest-floor samples¹² may also hold for solid-phase aluminosilicates in our mineral soils.

Following Cronan *et al.*¹², the exchange equilibrium $3\text{H}^+ + \text{AlX} = \text{H}_3\text{X} + \text{Al}^{3+}$ can be modelled as $[\text{Al}^{3+}]/[\text{H}^+]^3 = K_{\text{ex}}(\text{AlX}/\text{H}_3\text{X})^n = K_{\text{app}}$, where n , K_{ex} (exchangeable) and K_{app} (apparent) are constants, square brackets denote activities of aqueous ions and AlX and H_3X are their mole fractions on the organic sorbent. Taking negative logarithms, this expression becomes

$$p\text{Al} - 3p\text{H} = pK_{\text{ex}} - n \log(\text{AlX}/\text{H}_3\text{X}) = pK_{\text{app}}$$

The data in Fig. 3 indicate that pK_{app} increases with decreasing content of solid-phase aluminosilicates. Only the first equilibrium extracts of the podzol Bhs horizons (represented by diamonds in Fig. 3) were close to equilibrium with gibbsite.

In general, the KCl-extractable aluminium fraction is assumed to represent readily mobilizable, exchangeable aluminium¹⁴. However, our data show that in a strongly acidic environment the $\text{Na}_4\text{P}_2\text{O}_7$ -extractable aluminium fraction is a better estimate of readily mobilizable aluminium, exchanged against H^+ .

In the field, 1–2% of organic aluminium is removed annually from the surface layers of the Dutch soils (Table 1). Note that the current rate of mobilization of aluminium from the uppermost soil layers is lowest where the content of soil organic aluminium is lowest. Based on the content of soil organic aluminium, the mobilization rates of aluminium are potentially high in the spodic B horizons. However, the availability of readily mobilizable aluminium is associated with a rapid acid consumption and consequently an increase in solution pH, which restricts further aluminium mobilization. Soil solutions in Bhs horizons are generally close to equilibrium with gibbsite. In the Dutch podzols the current depletion of aluminosilicates is most pronounced in the upper few centimetres of the Bhs horizon, which was formed by precipitation of organic aluminium in the course of soil development. In such soils the

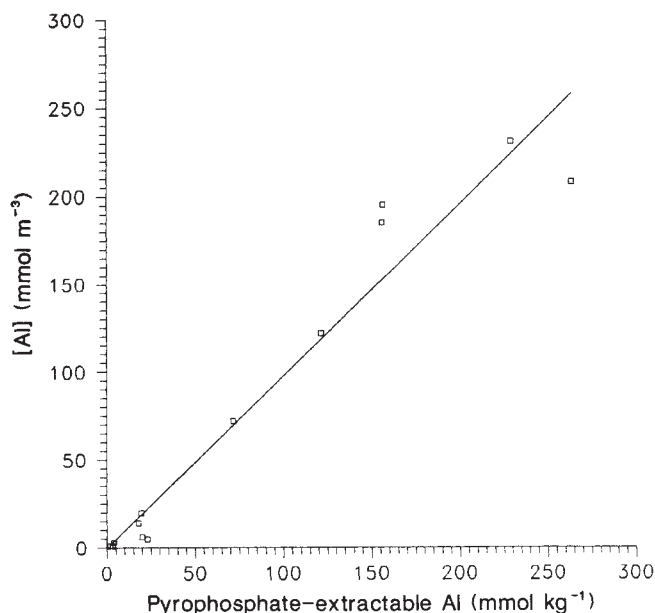


Fig. 2 Equilibrium aluminium concentrations in individual HCl leachates, after 10, 25 and 50 acid additions, as a function of $\text{Na}_4\text{P}_2\text{O}_7$ -extractable aluminium. Data refer to all six soil horizons. Equilibrium concentrations of aluminium ($[\text{Al}]$) are significantly correlated ($R^2 = 0.9466$) with $\text{Na}_4\text{P}_2\text{O}_7$ -extractable soil aluminium (Al_{pyro}): $[\text{Al}] = -0.60 + 0.98 \times \text{Al}_{\text{pyro}}$, where $[\text{Al}]$ is in mmol m^{-3} and Al_{pyro} is in mmol kg^{-1} .

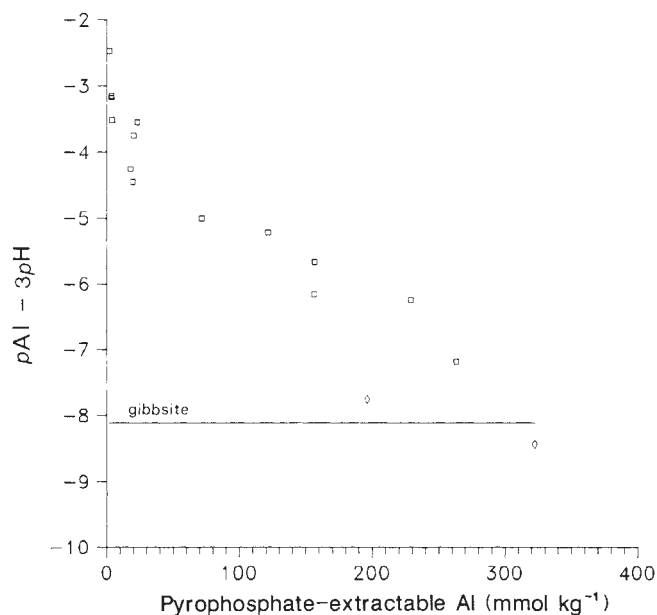


Fig. 3 Values for $pAl - 3pH$ in equilibrium leachates as a function of $Na_2P_2O_7$ -extractable aluminium. Leachates are from all soil samples after 10, 25 and 50 HCl additions (squares). For the samples of the two podzol Bhs horizons values of the first leachate are also given (diamonds). Aluminium activities were calculated using the chemical equilibrium program ALCHEMI (ref. 18). For gibbsite we used $pK = -8.11$ (ref. 13).

introduction of acid rain produces a drastic change in the naturally occurring podzolization process.

The mobility of aluminium in the Hubbard Brook soils is significantly lower than in the Dutch soils (Table 1), because of higher soil-solution pH values, which is due, in turn, to a lower acid load and higher mobilization rates of base cations¹⁵.

The current rapid depletion of organic aluminium in the rooting zone of many acid, sandy soils is irreversible on a timescale of decades or centuries. Because aqueous aluminium concentrations are in equilibrium with solid-phase aluminosilicates, depletion upon continued acid deposition will inevitably lead to a reduction in dissolved aluminium concentrations. At the same time, soil-solution pH values, which are currently buffered in the pH range 3.5–4.2 by dissolution of aluminium³ (J.M. *et al.*, preprint), may drop to values just below 3. Such dramatic changes in the soil solution chemistry of the rooting zone may have significant ecological consequences. Therefore, depletion of solid-phase organic aluminium and the associated changes in the chemical composition of soil solutions should be incorporated in current soil-acidification models.

We thank the Dutch State Forest Service and Mr W. A. M. Rijken for making available the research sites. This study was supported in part by the EEC and by the Netherlands Technology Foundation (STW).

Received 11 August; accepted 22 November 1988.

1. Van Breemen, N., Driscoll, C. T. & Mulder, J. *Nature* **307**, 599–604 (1984).
2. Khanna, P. K., Prenzel, J., Meiwes, K. J., Ulrich, B. & Matzner, E. *Soil Sci. Soc. Am. J.* **51**, 446–452 (1987).
3. Mulder, J., van Grinsven, J. J. M. & van Breemen, N. *Soil Sci. Soc. Am. J.* **51**, 1640–1646 (1987).
4. Andersson, F. & Kelly, J. M. (eds) *Aluminium Toxicity to Trees* (Swedish University of Agricultural Sciences, Uppsala, Sweden, 1984).
5. Pavan, M. H. & Bingham, F. T. *Soil Sci. Soc. Am. J.* **46**, 993–997 (1982).
6. Wright, R. J., Baligar, V. C. & Wright, S. F. *Soil Sci.* **144**, 224–232 (1987).
7. Ulrich, B., Mayer, R. & Khanna, P. K. *Soil Sci.* **130**, 193–199 (1980).
8. Richburg, J. S. & Adams, F. *Soil Sci. Soc. Am. Proc.* **34**, 728–734 (1970).
9. Bloom, P. R., McBride, M. B. & Weaver, R. M. *Soil Sci. Soc. Am. J.* **43**, 488–493 (1979).
10. David, M. B. & Driscoll, C. T. *Geoderma* **33**, 297–318 (1984).
11. Driscoll, C. T., van Breemen, N. & Mulder, J. *Soil Sci. Soc. Am. J.* **49**, 437–444 (1985).

12. Cronan, C. S., Walker, W. J. & Bloom, P. R. *Nature* **324**, 140–143 (1986).
13. May, H. M., Helmke, P. A. & Jackson, M. L. *Geochim. cosmochim. Acta* **43**, 861–868 (1979).
14. Kamprath, E. J. *Soil Sci. Soc. Am. Proc.* **34**, 352–354 (1970).
15. Mulder, J. & van Breemen, N. in *Effects of Atmospheric Pollutants on Forests, Wetlands and Agricultural Ecosystems* (eds Hutchinson, T. C. & Meema, K. M.) 361–376 (Springer, Berlin, 1987).
16. USDA *Soil Survey Laboratory Methods and Procedures for Collecting Soil Samples* (USDA, Washington DC, 1972).
17. Parfitt, R. L. & Childs, C. W. *Aust. J. Soil Res.* **26**, 121–144 (1988).
18. Schecher, D. W. & Driscoll, C. T. *Wat. Resour. Res.* **23**, 525–534 (1987).

A new picture of the Moho under the western Alps

ECORS-CROP Deep Seismic Sounding Group*

The Moho—the boundary between crust and mantle first discovered by A. Mohorovicic¹—is the main seismic marker in the continental lithosphere. The seismic nature of this interface, in terms of its position, topography, smoothness and continuity, may preserve imprints of the regional geodynamic evolution of the lithosphere. Here we report the results of a wide-angle seismic profiling experiment across the western Alps, which allows us to draw a cross-section of the Moho across this mountain belt. A tight sampling of this deep reflector shows abrupt changes in its depth and dip. The root zone of the chain (the zone of maximum crustal thickness) is well defined, with a 55-km-deep crust–mantle boundary. The Moho under the western Po plain is also clearly seen, and a shallow reflective structure is mapped under the Briançonnais zone in the 25–30 km depth range. This structure, if interpreted as lower-crustal or upper-mantle material, would support the hypothesis of a flaking of the lithosphere under the western Alps^{2,3}.

To prepare the layout of the vertical reflection line through the western Alps that formed the basis of the ECORS-CROP project^{4,5} in 1986, a preliminary series of experiments⁶ was conducted which aimed at mapping very deep interfaces along a cross-section extending from Grenoble (France) to the Po Plain (Italy). One-ton charges were detonated at five places (Fig. 1) and part of the recording array—low-frequency geophones and autonomous recorders—was deployed along fan profiles, a method which had already proved successful in other orogenic belts^{7–9}. The fan radii were chosen so as to coincide with the theoretical maximum amplitude of reflections, corresponding to total reflection, at and beyond a critical distance. The Moho depth was assumed to be ~40 km, with an increase from west to east; the fan radii therefore ranged from 90 to 130 km. Because the velocity is assumed to be constant within the crust, and because each shot-station line is along the local strike of the Alps, no apparent dip of the reflector is expected. Each mirror point (triangles in Fig. 1) can therefore be plotted half-way between the shot-point and the station. With an average station spacing of 4 km along the fans, the reflectors are thus sampled every 2 km.

The fan data were processed to construct two composite cross-sections of the Alpine chain, showing the topography of deep reflectors (Figs 2 and 3). Figure 2a, corresponding to shots A, B and C, begins 30 km west of the Belledonne External Crystalline Massif (ECM) and extends through the Briançonnais and Piemontese zones to the east of the Dora Maira Internal Crystalline Massif (ICM); Fig. 2b (shots D and LW) extends from the Gran Paradiso ICM to the Po Plain, intersecting the Sesia-Lanzo zone and the Canavese line, which marks the western limit of the chain. It should be understood that these

* A. Hirn & S. Nadir, Institut de Physique du Globe, 4 place Jussieu, Paris, France; F. Thouvenot, Laboratoire de Géophysique Interne et Tectonophysique, Observatoire de Grenoble, IRIGM, 38041 Grenoble, France (contact address); R. Nicolich & G. Pellis, Istituto di Miniere e Geofisica Applicata, Università di Trieste, I-34127 Trieste, Italy; S. Scarascia & I. Tabacco, Istituto di Geofisica della Litosfera and Istituto Fisica Terrestre, Milano, Italy; F. Castellano, Osservatorio Vesuviano, Napoli, Italy; F. Merlanti, Istituto di Geofisica, Università di Genova, Italy.

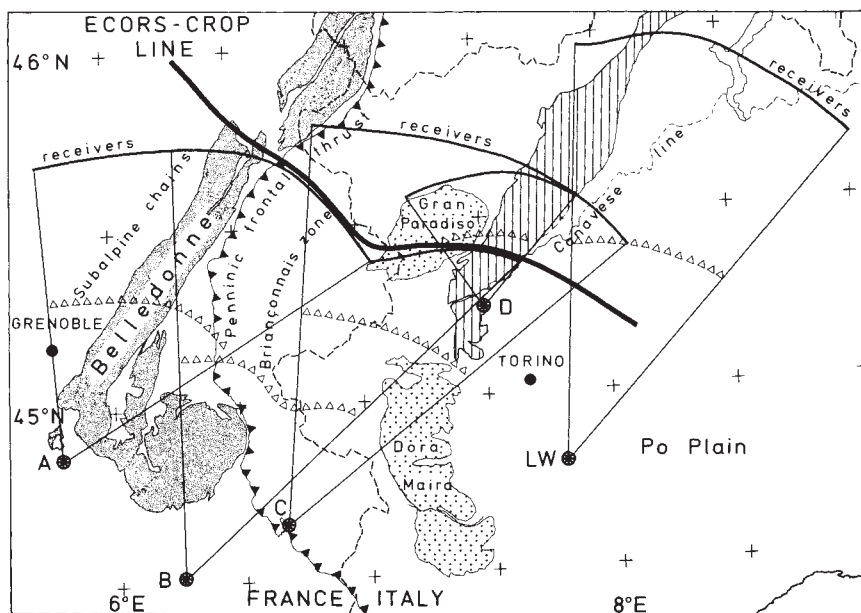


Fig. 1 Position map of the fan layout, showing the 5 shot-points A, B, C, D and LW. Open triangles indicate reflection points used to construct Fig. 2. Shaded area represents external crystalline massifs; dotted area represents internal crystalline massifs; hatched area represents the Sesia-Lanzo unit.

cross-sections combine fans with unavoidable offsets between them; for example, in Fig. 1, mirror points for shot D are shifted northwards by about 50 km relative to those for shot B. For this reason we chose to separate the two northern fans (D and LW) from the three southern ones, to keep a minimum of continuity within a composite cross-section. According to the Bouguer gravity map¹⁰, the deep structures can be assumed to retain a cylindrical geometry for at least the 30-km extension across the section considered here.

On the seismic sections, reflections are picked using a maximum-amplitude criterion together with a correlation between traces. Around the critical distance, the wavelet reflected from the Moho is known to be very energetic, so that it can be identified with the maximum-amplitude signal. The sharpness of the onset of the reflected wavelet varies throughout the section (compare, for example, the two ends of the section in Fig. 2a), probably because of a variation in the seismic response of the reflector. Large (~20-km) variations in depth of the Moho marker are evident, and dips reaching 20° and extending over distances of tens of kilometres are observed. In the ECM area (Figs 2a and 3), the changes in depth and dip do not occur smoothly, as if by continuous flexure: from being almost horizontal at a depth of 37 km under the sub-alpine massifs and

the ECMs, the Moho dips abruptly from 37 km to 55 km at their eastern edge, where the present-day maximum surface relief is found. Towards the innermost parts, under the Piedmontese zone and the Dora Maira ICM, it remains at a more or less constant, deep level of 55 km. The conversion of reflection times to depths depends on the mean crustal velocity, which is here taken to have the value 6.25 km s^{-1} ; from previous seismic experiments¹¹⁻¹⁵, this value is known to be rather constant over this region so that we can assume that the depths thus obtained are accurate to better than 5%.

This deep Moho is overlaid by another seismic marker with high reflectivity (Fig. 2a, shot C). The maximum amplitude of the signals from the latter occur at depths of between 25 and 30 km, although it is difficult to determine accurate arrival times for these. Common experience in deep seismic profiling suggests that this reflective zone should be associated with a velocity contrast of more than 1 km s^{-1} , so that it corresponds either to the top of lower-crustal material or to a crust-mantle boundary. Situated at shallow depth and extending westwards beneath the Piedmontese zone and the ICMs as far as the Briançonnais zone, this reflector can be detected only when shooting from C. It acts as a mask which prevents the seismic energy from penetrating deeper, so that from shot-point C no deep Moho

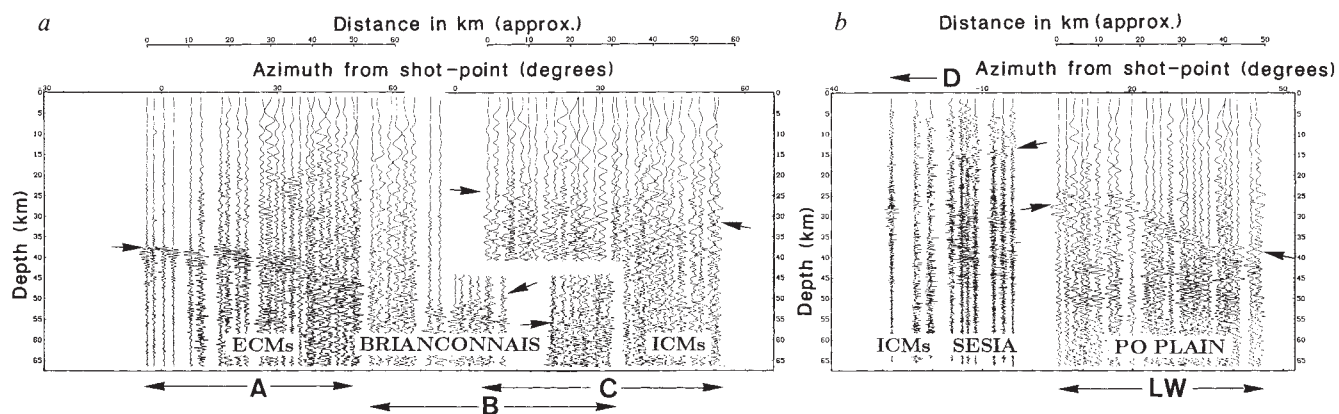
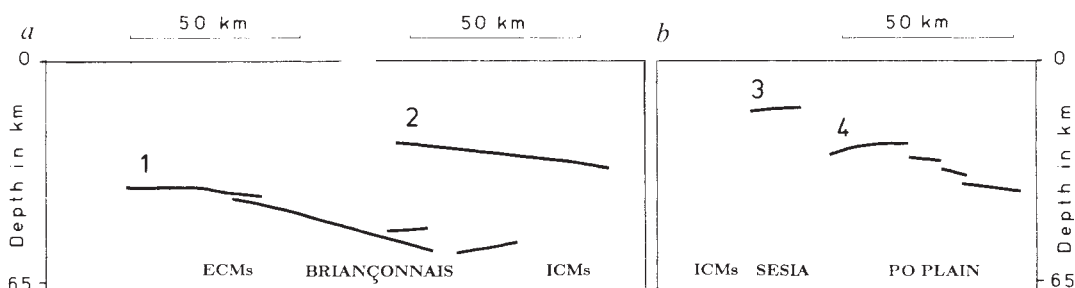


Fig. 2 Two composite cross-sections of the western Alps. *a*, The three fan profiles from shot-points A, B and C (see position on Fig. 1) are combined as a single record-section extending from the Chartreuse sub-alpine massif (far left) to the Dora Maira massif (far right). *b*, This shorter section combines results from shot-points D and LW and extends from the Gran Paradiso massif (far left) to the Po plain (far right). In each trace, the maximum-amplitude signal corresponds to reflection from the Moho. Earlier (shallower) energy arrivals may be ascribed to intracrustal discontinuities. For shot D, recorded at the relatively short distance of 40 km, the late (deep) arrivals are principally transverse S waves rather than reflected waves. Depths are calculated assuming a mean crustal velocity of 6.25 km s^{-1} .

Fig. 3 Schematic line drawings of the profiles revealed by the corresponding seismograms in Fig. 2. (1) Autochthonous deep Moho; (2) Briançonnais reflector; (3) Ivrea body reflector; (4) hinterland (Po plain) Moho. In lithospheric flaking optics, features (2) and (3) trace the imbrication of anomalous upper mantle within the Alpine pile.



could be mapped. From shot-points A or B, however, the deep Moho can be reached, illustrating both the limited extent of this shallow unit and the importance of the shot/recording-array geometry in detecting such a feature.

The hinterland Moho (Fig. 2b), more than 35 km deep under the Po plain, where thick Tertiary-Quaternary sequences occur, seems to rise in a stepwise fashion to less than 25 km approaching the Canavese fault zone. Across this zone, there exists a gap in the seismic data, so that one cannot rule out the existence of another uprising of material with high seismic velocities. The Sesia zone is underlain by a strong reflector at a depth of only 13 km, which may correspond to a crust-mantle boundary. However, other deep reflectors in the same area, seen along a complementary longitudinal profile through Sesia, cast doubt on whether the hinterland Moho extends this far west and at such shallow depth; in a zone close to the Lanzo massif, where slices of lower-crustal and upper-mantle material have been brought up to the surface, reflective levels might be rather discontinuous. Nevertheless, the simplest structural scheme is to link the 13-km deep reflector discovered under the Sesia zone to the Po plain Moho further east.

From this wide-angle reflection profiling therefore emerges the first detailed picture of the geometry of the Moho beneath the Alps, incorporating even the innermost zones. Previous experiments¹¹⁻¹⁵ sampled the Moho too sporadically so that only smooth variations could be mapped. Close-ups of the sections presented here show that such a continuous geometry is not entirely correct, as was previously suggested by a teleseismic prospecting of lithospheric contrasts beneath the Alps¹⁶. This conclusion results mainly from the high spatial resolution made available by the arrangement of the fans, which provided a 2-km sampling interval of reflectors along the cross-sections.

The shallow reflective level that we have discovered under the Briançonnais zone has never before been mapped, although recently published models of Alpine orogenesis² have suggested such a feature. The presence of lower-crustal or upper-mantle material would indicate a tectonic thickening, resulting from previously adjacent eastern lithospheric segments having been thrust onto the western segment (the alternative, internal deformation of the sole crust, is unlikely). This would suggest flaking of the European lithosphere under the western Alps, which is discussed more thoroughly elsewhere⁵.

The Ivrea body, an anomalously shallow mantle unit discovered from gravity evidence in early geophysical experiments in the Alps¹¹, is sampled here from shot-point D, but our experiment is unable to define its exact geometry. It is much too shallow (13 km depth) to be connected directly to the Briançonnais reflector, 30 km under the ICMs. The anomalous upper mantle imbricated within the Alpine crust therefore consists of two distinct units, the Ivrea body and the new unit discovered here.

The fan layout of the experiment proved successful in detecting very deep reflections, and the ECORS-CROP vertical reflection line was able to provide information on upper- and mid-crustal reflectors. Uniting vertical reflection profiling with the more versatile wide-angle reflection method therefore provides

the sort of complementary information required to determine the deep structure of such complex orogens as the Alpine chain.

This reconnaissance project was funded by the ECORS (France) and CROP (Italy) national programmes. The participation of numerous field operators, as well as the support and willingness of local authorities at the shot-points, are acknowledged. We thank L. Jenatton for help with the artwork.

Received 6 June; accepted 29 November 1988.

1. Mohorovicic, A. *Jh. Met. Obs. Agram* **9**, 1-63 (1909).
2. Ménard, G. & Thouvenot, F. *Bull. Soc. géol. Fr.* **5**, 147-156 (1984).
3. Butler, R. W. H. *J. geol. Soc. Lond.* **143**, 857-873 (1986).
4. Bayer, R. *et al. C. R. Acad. Sci. Paris* **305**, 1461-1470 (1987).
5. ECORS-CROP Working Group (in preparation).
6. ECORS-CROP Deep Seismic Sounding Group, *Tectonophysics* (submitted).
7. Hirn, A. *et al. Geophys. Res. Lett.* **7**, 263-266 (1980).
8. Hirn, A. *et al. Nature* **307**, 23-25 (1984).
9. Hirn, A. *et al. Ann. Geophys.* **2**, 113-117 (1984).
10. Bayer, R. & Lanza, R. *Tectonophysics* (submitted).
11. Closs, H. & Labrousse, Y. *Mém. Coll. CNRS Paris, Ann. Géophys. Int.* **XII-2**, 241 (1963).
12. Choudhury, M., Giese, P. & de Visintini, G. *Boll. Geofis. teor. appl.* **XII/51-52**, 211-240 (1971).
13. Perrier, G. thesis, Univ. Paris (1973).
14. Giese, P. & Prodehl, C. *Explosion Seismology in Central Europe* (Springer, Heidelberg, 1976).
15. Miller, H., Müller, St. & Perrier, G. *Alpine Mediterranean Geodynamics, Am. Geophys. Un. Geodyn. Ser.* **7**, 175-203 (1982).
16. Hirn, A. *et al. Nature* **308**, 531-532 (1984).

Testing the completeness of earthquake catalogues and the hypothesis of self-similarity

Paul A. Rydelek & I. Selwyn Sacks

Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road N.W., Washington, DC 20015, USA

Self-similarity of the earthquake process is consistent with the observed linear b -value relation $\log(N) = a - bM$, where N is the number of earthquakes with magnitude M . Deviations from linearity are believed to be due to statistical fluctuations because of the scarcity of events at large M , or from incompleteness because of a detection threshold at small M . Above some magnitude level, all local events are detected, because they exceed the noise background on the seismogram. As magnitude decreases, however, events go undetected as the seismic signal approaches the noise background. Thus more of the smallest events should be logged at night than during the day, because the cultural noise sources and winds are diminished at night, resulting in presumably quieter seismograms. Here we use this day-to-night noise modulation to develop a completeness test for earthquake catalogues; catalogues that indicate no significant modulation are considered complete. We test three catalogues and show that the completeness magnitude can be different from the magnitude at which the b -value departs from linearity. In particular, catalogues that show significant deviations in linearity at small M but are otherwise complete, are at odds with the hypothesis of earthquake self-similarity.

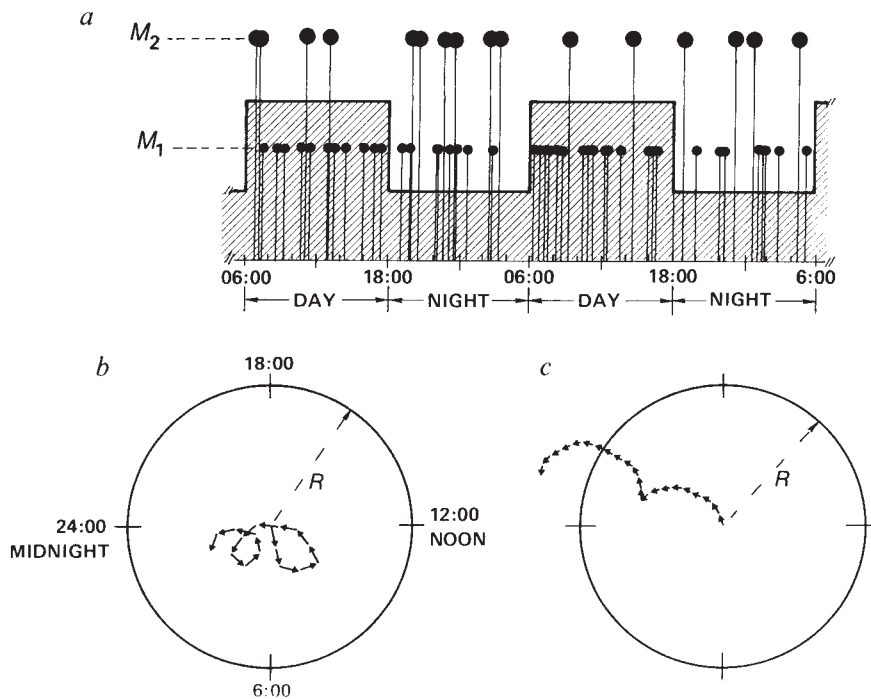


Fig. 1 *a*, An example of day-night noise modulation using a synthetic earthquake sequence of two different-magnitude events which follow Poisson distributions. The larger earthquakes (M_2) exceed the background noise (stippled area), so the phasor plot of these events (*b*) ($N = 16$, $R = 6.9$) shows no bias in the walkout. The smaller events (M_1) are detected only during the night, when the noise background is reduced, so these events show a significant phasor walkout (*c*) ($N = 50$, $R = 12.2$). The circles in *b* and *c* are drawn at the 95% confidence level for an equivalent random walk (Brownian motion). The Poisson rate for the smaller events is 1 event per hour and the b -value is 0.5.

Our method of testing for completeness follows one originally proposed by Schuster¹. The test makes two assumptions about the earthquake process and the seismographic station, namely (1) that earthquakes, at any magnitude level, follow a Poisson distribution, and (2) that the daytime noise level is increased relative to night-time because of cultural activity and wind noise.

The randomly distributed (Poisson) nature of earthquake sequences (with aftershocks removed) is well documented in studies of seismic catalogues² but exceptions have been found, namely in a few cases that show modulations of the seismicity from tidal stresses of the Sun and the Moon^{3,4}, that is, through tidal triggering of earthquakes. A spectrum of the tidal forcing function, however, has negligible power at a period of exactly 24 hours⁵, which is the period of the clock used in our test of completeness described below. The only known noise sources with a 24-hour periodicity are cultural and solar thermal effects (such as surface temperatures and winds); thus any daily modulations in the seismicity must be due to these sources. When observed, a modulation implies that some events at the detection threshold are missed during the half-cycle of increased noise, resulting in an incomplete catalogue. Our method of analysis, but with clocks tuned instead to the major tidal components, has been used successfully in studies of tidal triggering of earthquakes³. We have ignored the remote possibility of thermally triggered earthquakes, but note that the phase in Fig. 1 could be diagnostic.

Within a given magnitude level, the origin time of a seismic event is used to form a phase (or hour) angle on the local 24-hour clock; thus the catalogue of earthquake times is mapped into a set of earthquake phases. Each phase angle is assigned unit amplitude and then a phasor sum is performed over the set of individual unit phasors. The resulting phasor sum or 'walkout', R , is compared with that expected from a random process. For a set of random phases (that is, earthquakes that follow a Poisson process) the probability of obtaining a walkout larger than R is given by $P_R = \exp(-R^2/N)$. We reject the null hypothesis of randomly distributed earthquake phases when P_R is below some confidence level ($1 - P_R$). Therefore, if the observed R exceeds some critical radius (a reasonable value is $1.73\sqrt{N}$, corresponding to a 95% confidence level), then the catalogue contains a significant bias in the day-to-night record-

ing of earthquakes in the given magnitude interval and should be classified as incomplete at (and below) this magnitude. The sensitivity of our test is dependent on the number of events. For example, we can detect a 3% incompleteness (worst case) in a 10,000-event catalogue, increasing to 9% incompleteness for 1,000 events.

A simple example of day-night noise modulation is shown in Fig. 1. In this example, all the larger-magnitude earthquakes would be detected as they are clearly above the noise at any time of the day. The phasor walkout formed from this set of earthquakes grows as $\sqrt{N}$ and resembles Brownian motion because the phase angles are random, having been determined from earthquakes that have a Poisson distribution. The smaller-magnitude earthquakes are also Poisson-distributed but because more are detected when the noise is reduced (that is, during local night) they tend to form a non-random pattern with average alignment biased towards night-time hours. In this example, our test indicates that the larger-magnitude events are completely catalogued but the smaller ones are not. Of course, variations in noise levels in real catalogues are never as clear-cut as shown in our example, but the persistent bias introduced by any day-night noise modulation should eventually be detectable in catalogues of longer duration. As soon as the modulation becomes significant, the catalogue is shown to be incomplete. Intrinsic in the test is the confirmation that the modulation exists, because it becomes visible at lower magnitudes. The test provides an objective answer to the question of completeness, makes no assumptions about the linearity of the b -value curve, and allows one to set a confidence level on catalogue reliability to answer further questions regarding the underlying nature of the earthquake process.

Other problems arising in catalogues are the result of bursts of earthquake activity, such as swarms and aftershocks, which produce sections of highly non-random, and in worst cases almost linear, phasor walkout. There are algorithms for identifying (and removing) aftershocks², but we choose simply to ignore all shocks occurring more frequently than at twice the average hourly earthquake rate for the catalogue (again within a given magnitude interval).

We demonstrate our completeness test with catalogues of crustal earthquakes (0–30 km deep) obtained from three

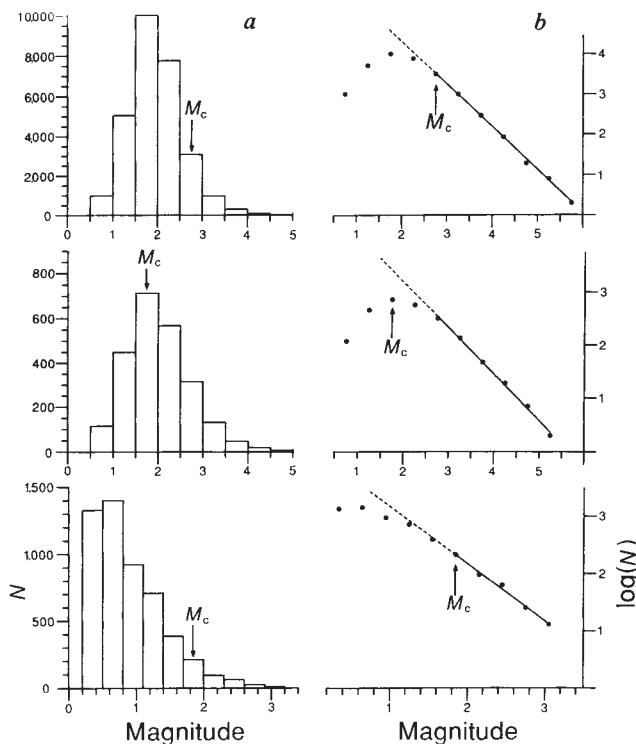


Fig. 2 Plots of a , the number of events N , and b , $\log(N)$, within given magnitude intervals for catalogues HAW (a), KMU (b) and POZ (c). The b -values are, respectively, 1.06, 0.87 and 1.00, and were determined using the points fitted with the solid straight line; the dashed line is an extrapolation using $\log(N) = a - bM$. Arrows mark the smallest magnitude at which a catalogue is complete, based on the phasor walkouts in Fig. 3.

different seismic arrays: (HAW) Island of Hawaii, coverage of ~ 80 km radius, 41 stations, 1977–1985; (KMU) southeast corner of Hokkaido, Japan, ~ 50 km radius, ~ 9 stations, 1976–1987; and (POZ) Phlegraean Fields near Pozzuoli, Italy, ~ 3 km radius, 22 stations, 1983–1984. Plots of the number of events versus magnitude are shown in Fig. 2 for these catalogues. The difference in the peak position and shape of the curves between the larger arrays HAW and KMU, and the smaller, concentrated array POZ is readily apparent (Fig. 2a), but b -values are still near unity (Fig. 2b), in general agreement with the concept of self-similarity. Notice also that catalogue POZ is divided into smaller magnitude intervals because of the increased sensitivity but limited-magnitude bandwidth of this array. The phasor walkouts as a function of magnitude for the three catalogues are shown in Fig. 3 for the same magnitude intervals used to construct Fig. 2. The smallest magnitude interval at which each of the catalogues passes the completeness test (95%) is labelled as M_c . We are confident that the data set at (and above) M_c is complete but we make no claims about the degree of incompleteness of the data below this point.

Catalogue HAW shows a departure from linearity in the b -value plot at $\sim 1/2$ magnitude below M_c . We assume that this non-linearity is caused by the loss of events below the threshold, rather than having the same cause as the violation of the linear relation seen for larger magnitudes.

In POZ, an obvious non-linear departure starts at ~ 1 magnitude below M_c and rolls off rather gradually. This catalogue is collected from a relatively small region compared with KMU and HAW and is therefore susceptible to very localized noise sources such as commuter traffic on a road during certain times of the day. Although this selective increase in background noise may have a small effect on the b -value curve, say a $\sim 10\%$ deficit,

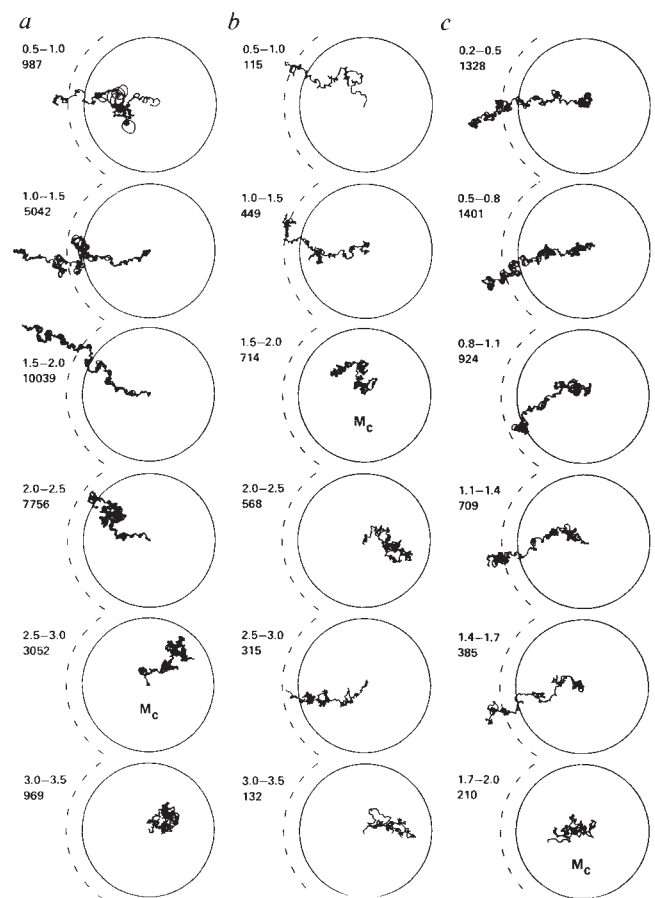


Fig. 3 Phasor sum plots for the catalogues HAW (a), KMU (b) and POZ (c). The numbers in each frame are the magnitude intervals and the numbers of events. The plots are time-oriented as in Fig. 1b, and scaled such that the confidence levels are drawn with constant radii (solid = 95%; dashed = 99% confidence level). A catalogue whose walkout does not span a chosen confidence level is considered complete at that level; thus M_c is the smallest magnitude of completeness (95%) for each catalogue in this figure.

it is certainly within the range of detectability by our completeness test. POZ exemplifies a catalogue that does not show significant departures from linearity even though it fails the completeness test. In fact, when the number of events peaks, our test attains maximum sensitivity, in contrast to a b -value plot which has minimum sensitivity at peak N .

The KMU catalogue shows a highly significant fall-off in the number of smaller events even though the catalogue is apparently complete. Taylor and coworkers^{6,7} reached the same conclusion in a recent study of this catalogue. We believe that this points to an important deviation in the common lore of linear b -value curves and the idea of the self-similar earthquake process.

We thank the following (and their colleagues) for kindly allowing us to use their catalogues: (HAW) R. Koyanagi, Hawaiian Volcano Observatory, (KMU) T. Takanami, Hokkaido University, and (POZ) R. Scarpa, University of Rome.

Received 29 September; accepted 23 November 1988.

- Schuster, A. *Proc. R. Soc. Lond.* **61**, 455–465 (1897).
- Gardner, J. K. & Knopoff, L. *Bull. seismol. Soc. Am.* **64**, 1363–1367 (1974).
- Young, D. & Zürn, W. *J. Geophys.* **45**, 171–182 (1979).
- McNutt, S. R. & Beavan, R. J. *Nature* **294**, 615–618 (1981).
- Melchior, P. *The Tides and the Planet Earth* (Pergamon, Oxford, 1978).
- Taylor, D. thesis, Virginia Polytechnic Inst., Blacksburg, Virginia (1988).
- Taylor, D., Snoke, J. A., Sacks, I. S. & Takanami, T. *Seismol. Res. Lett.* **58**, 105 (1987).

Distribution of seamounts in the North Atlantic

David Epp* & N. Christian Smoot†

* National Science Foundation, 1800 G Street NW, Washington, DC 20550, USA

† Bathymetry Division, US Naval Oceanographic Office, NSTL, Mississippi 39522, USA

The paucity of multi-narrow-beam bathymetric data in the ocean basins has meant that previous studies^{1–8} of seamount distribution and size have relied on wide-beam echo sounder data and small amounts of Seabeam multi-narrow-beam data. Since 1967, however, the US Naval Oceanographic Office has been collecting multi-narrow-beam bathymetric data^{9–14}. Here we use an extensive collection of such data to locate over 800 seamounts in the North Atlantic. We find that ridge crest-generated seamounts are formed only between Iceland and the Hayes fracture zone, where there are contributions of plume material to the mid-ocean ridge plumbing system. It appears that in the slow-spreading North Atlantic an injection of plume material into the ridge plumbing system is necessary for the formation of seamounts at the ridge crest.

We use the term seamount for all circular or elliptical features of volcanic origin, regardless of size, including those where flanking rift zones or slumping alter the basic circular or elliptical shape. The multi-beam data we studied are contoured at 10-fathom intervals (1 fathom = 6 feet). In some areas, such as the rift valley of the northern Reykjanes ridge¹⁵, we were able to locate seamounts that had a relief of the order of 20–30 fathoms. Because of the rough block-faulted nature of the ridge, however, it is impossible, on the ridge flanks, to separate seamounts this small from the general ridge-flank tectonic fabric. To be certain that we counted only volcanic features, we have limited our study to seamounts that have a relief greater than 10 fathoms. Although there are undoubtedly smaller volcanoes that we did not count, the smaller seamounts, where detectable, have the same general distribution as those higher than 50 fathoms.

Figure 1 shows the location of seamounts in the North Atlantic with a relief greater than 50 fathoms. Excluding the Lesser Antilles, there are 810 seamounts. The blank areas reflect the lack of seamounts; the general distribution pattern is unlikely to be affected by the lack of data. We estimate that the total number of seamounts with heights greater than 50 fathoms is unlikely to be underestimated by more than 10% as a result of incomplete data coverage. Sediment thickness in the North Atlantic is highly variable¹⁶ and it is likely that smaller volcanoes have been buried. For example, in the area between the New England and Corner seamounts there are a number of small seamounts that are probably the tops of larger seamounts buried by sediments from the Sohms abyssal plain. The pattern of sediment distribution in the North Atlantic does not, however, correspond to the distribution of seamounts shown in Fig. 1, indicating that the general seamount distribution pattern is unlikely to be affected by sediment burial.

Morgan^{17,18} and Duncan¹⁹ have described the traces of the major hotspots in the North Atlantic. Figure 1 shows the distribution of seamounts in these traces, in addition to several smaller traces. Madeira and Canary hotspots generated the Milne and Newfoundland seamount groups, respectively, before crossing to the east side of the Mid-Atlantic Ridge (MAR). The Canary hotspot is probably responsible for the cluster of seamounts located at 38°–41° N and 15°–18° W. Both of these hotspots passed through the Madeira–Tore Rise and Horseshoe seamount areas¹⁹. In addition, the Africa/Eurasia plate boundary passes through the Horseshoe seamount area, so it is not clear which of the seamounts in this area were generated directly by the hotspots and which might be due to hotspot interaction with

the plate boundary. The trace of the Azores hotspot is uncertain. According to Morgan¹⁷ the hotspot would have been near Coruna 'seamount' at about 60 Myr ago and would have followed the Azores–Biscay rise to its present position. (Coruna is located at 44° N 14° W, but it does not appear to be volcanic and is not plotted in Fig. 1.) The trace was undoubtedly affected by interactions with the MAR and the Iberia–Africa/Eurasia plate boundary. Great Meteor hotspot generated the New England and Corner seamounts before crossing the MAR and generating the seamounts in the vicinity of Cruiser and Great Meteor. The hotspot was near the ridge crest during the formation of the Corner seamounts, ~75 Myr ago, and remained near the crest until ~38 Myr ago, when it crossed to the east side of the ridge between the Oceanographer and Hayes fracture zones¹⁹. The group of seamounts between the New England seamounts and Bermuda can be attributed to Verde hotspot¹⁹, but it is not clear if the three small groups located along 26°–27° N between 20° and 28° W are also due to Verde. Also, the origin of the three clusters west of the Cape Verde islands is unclear. The Researcher seamounts and the Barracuda/Researcher ridge form chains of volcanoes that are co-polar with the other hotspot traces in the western North Atlantic and are thus probably due to weak hotspots. The North Brazilian ridge is a line of seamounts in the Ceara abyssal plain that also may be a hotspot trail and may be related to either the Bathymetrists or Sierra Leone hotspots. The Charcot 'seamounts', located along 45° N between 10 and 15° W, do not appear to be volcanic in origin and are probably an expression of the extinct Iberia–Africa/Eurasia plate boundary.

The interaction of hotspots with mid-ocean ridge (MOR) systems has been discussed extensively^{20–26}. Morgan²⁷ suggested that the hotspot did not have to be directly under the MOR for this interaction to occur; conduits for the flow of plume material to the MOR plumbing systems could be maintained as the MOR moved away from the hotspot. Schilling^{28,29} has elaborated on this model and, on the basis of extensive sampling of the crest of the MAR, proposed that along-ridge geochemical gradients about ridge-centred hotspots can be attributed to incompatible-element-rich magmas which rise as plumes or blobs beneath the hotspots and mix with the incompatible-element-depleted asthenosphere that normally feeds the spreading ridge. In the North Atlantic, Schilling²⁹ identifies hotspot contributions from Jan Mayen, Iceland, 45° N, Azores, 35° N and 14° N. The K₂O, La/Ti and La/Sm anomalies associated with Iceland extend south of Iceland to about 60° N. The strontium anomaly reaches a minimum at 60° and appears to increase toward the Charlie–Gibbs fracture zone. The Iceland helium anomaly, however, extends south to the Charlie–Gibbs fracture zone (see Schilling²⁹, Fig. 3 and Plate 8B). There is a narrow geochemical anomaly associated with the Charlie–Gibbs fracture zone. Between the Charlie–Gibbs and Kurchatov fracture zones there is a broad geochemical anomaly centred at 45° N. There appears to be a narrower superimposed anomaly at 43° N. These anomalies cannot be attributed to known hotspots. If they are due to separate hotspots, then these hotspots did not generate seamount chains or other bathymetric expressions. The geochemical anomalies caused by the Azores hotspot extend between the Kurchatov and Hayes fracture zones. There is a small contribution at 35° N between the Pico/East Azores and Hayes fracture zones that probably comes from Great Meteor hotspot³⁰. South of the Hayes fracture zone there is no plume component except for a small anomaly at 15° N, for which Barracuda/Researcher hotspot may be responsible.

There is a first-order correlation between the distribution of seamounts along the MAR and the hotspot contribution to MAR magmas. Seamounts are formed only between Iceland and the Hayes fracture zone where there are contributions of plume magma to the MAR plumbing system. South of the Hayes fracture zone there are no ridge crest-generated (RCG) seamounts and only the small geochemical anomaly at 14° N. The

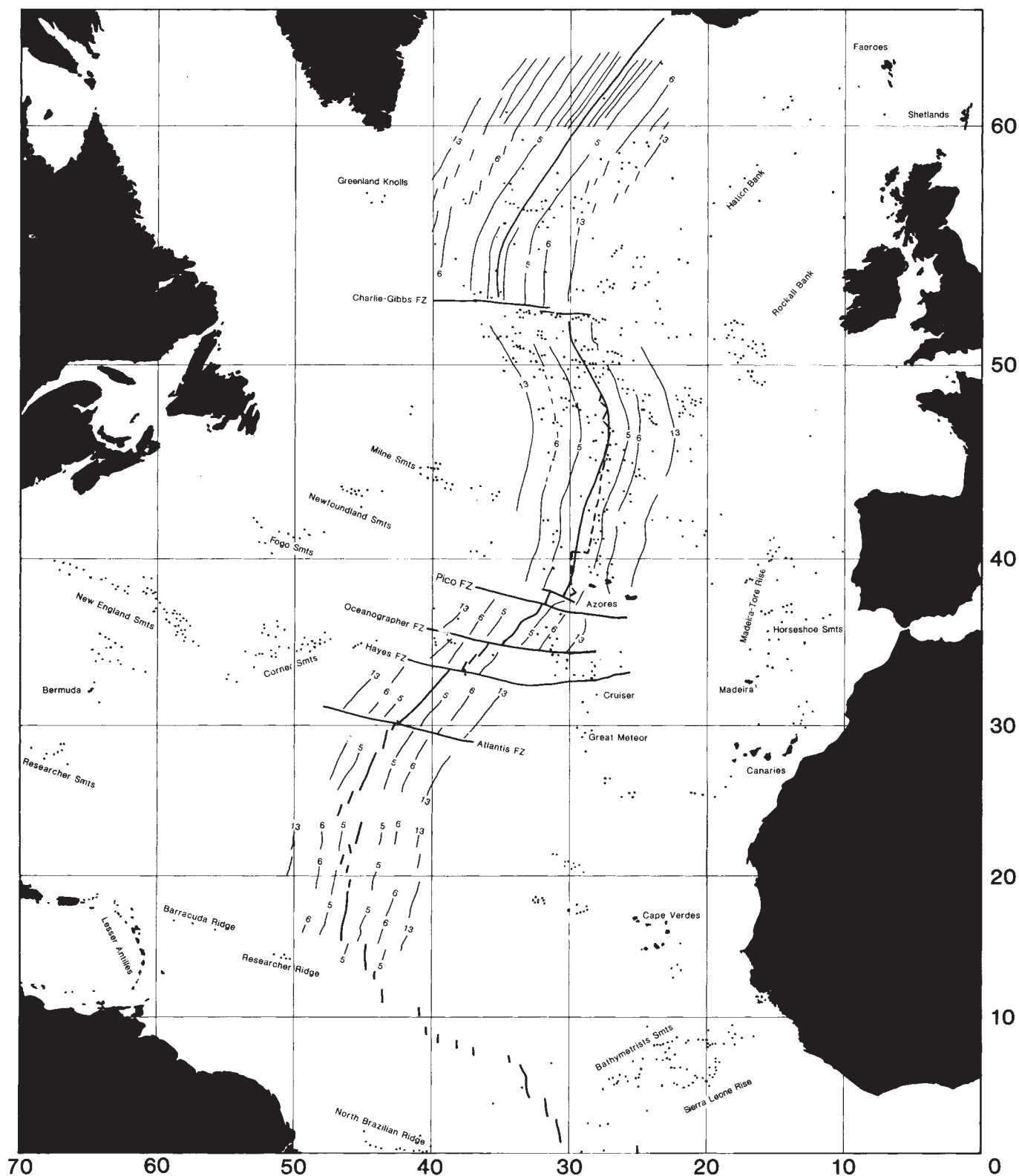


Fig. 1 The locations of seamounts in the North Atlantic with heights greater than 50 fathoms (shown as dots). Magnetic lineations are from S. Cande (personal communication). The dashed line shows the location of the topographic high between the Charlie-Gibbs fracture zone and the East Azores fracture zone; it does not correspond to anomaly zero.

latter is relatively recent³¹ and it may be that the Barracuda/Researcher hotspot has contributed only enough material to the plumbing system to affect the geochemistry but not enough to generate seamounts. The detailed relationships between the distribution of seamounts and plume contributions to the MAR plumbing system are more complex. In addition to the variation

in seamount numbers along the MAR, it is obvious from Fig. 1 that seamount generation varies with time and is not always symmetrical about the ridge axis. Figure 2 shows the number of seamounts for every one degree of latitude in 0–10, 10–20 and 20–35 Myr time windows (corresponding to magnetic anomalies 1, 5, 6 and 13) for the east and west sides of the

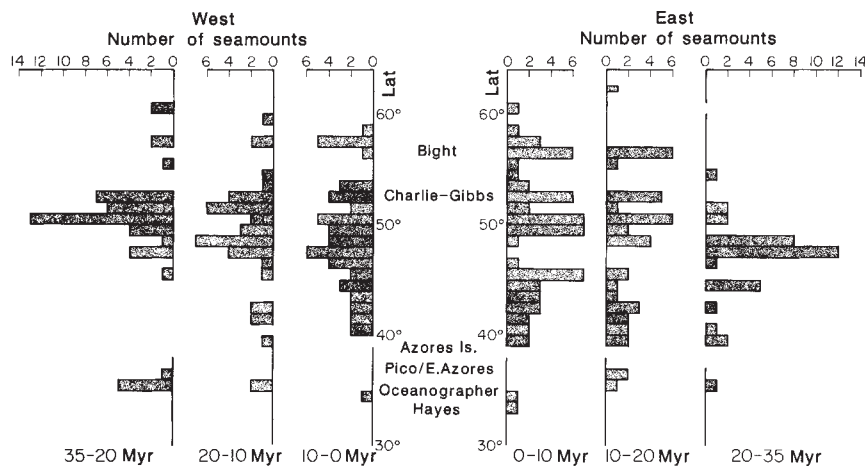


Fig. 2 The number of seamounts plotted against latitude in 0–10, 10–20 and 20–35 Myr time windows (corresponding to magnetic anomalies 1, 5, 6 and 13) for the east and west sides of the Mid-Atlantic Ridge. The latitudes of fracture zones and the Azores islands are indicated. Where the base line of the graphs is omitted there is insufficient data to provide an accurate count of seamount numbers.

MAR. The greatest concentration of seamounts occurs between the Charlie-Gibbs and the latitude of the Azores islands. This presumably reflects the strong contribution of plume material to this section of the MAR by the 45° N and Azores hotspots. The distribution of seamounts on the west side of the ridge suggests that these hotspots started contributing material to the plumbing systems at ~45 Myr in the section of ridge just south of the Charlie-Gibbs, and that the interaction is no older than ~20 Myr at the present latitude of the Azores. On the east side of the ridge the seamount distribution indicates an older and more complex interaction. There is a significant cluster of seamounts on crust older than anomaly 13 between 47° and 49° N, an absence of seamounts between the ridge crest and anomaly 6 at 46° to about 49° N, and a large cluster just west of the Ireland continental shelf at 49° to 52° N. This complexity in seamount distribution is most probably related to the reorganization in plate boundaries in this area and the interaction between the ridge plumbing system and the hotspots. Between ~60 and 35 Myr, King's trough, the Azores-Biscay rise and the Charcot 'seamounts' formed the plate boundary between the Iberia-Africa and Eurasia plates^{32,33}. The Azores hotspot was moving along the Azores-Biscay rise during this time¹⁷; however, the location (or existence) of the 45° N hotspot during this time is uncertain as it did not leave a bathymetric trace.

The seamounts between the Azores and the Hayes fracture zones may be due to the plume contribution from Azores hotspot or from Great Meteor hotspot. Morgan¹⁸ suggested that the group of seamounts between Cruiser and Great Meteor were formed directly by Great Meteor hotspot and that the cluster of seamounts between Cruiser and the Oceanographer fracture zone was formed by the interaction of the hotspot with the MAR. Great Meteor hotspot is nearly 900 km from this section of ridge crest²⁸, so its contribution to the MAR plumbing system is probably small. The geochemical anomaly associated with the Azores, on the other hand, extends past the Pico/East Azores fracture zone to the Hayes. The offset across the Pico/East Azores fracture zone decreases significantly between anomalies 6 and 5. In the vicinity of the Azores the ridge crest appears to be 'pulled' toward the Azores (as indicated by the dashed line in Fig. 1), and although there is a distinct bathymetric expression of the fracture zone east and west of the ridge crest, there is no distinct bathymetric offset of the present ridge crest. Thus, there may be no blockage of flow at the Pico/East Azores fracture zone. On the other hand, the offset on the Oceanographer fracture zone is 120 km and the offset on the Hayes is 80 km³⁴. It is not clear why material flows past the Oceanographer but not past the Hayes, unless the volume of flow is nearly depleted by the time it gets to the Hayes.

With the exception of the helium anomaly, and perhaps the strontium anomaly, the geochemical anomalies associated with

Iceland hotspot do not extend south of 60° N²⁹. There is, however, no offset in the MAR at this latitude that would indicate a blockage of flow. There are very few seamounts along the Reykjanes ridge between Iceland and 60° N whereas there are a significant number between 60° N and the Charlie-Gibbs fracture zone. Seamounts have been generated along this section of the MAR for the past 35 Myr, although the seamounts around the periphery of Rockall bank and the cluster between the southwest corner of Rockall and the Charlie-Gibbs may have been formed at the ridge crest before 35 Myr. The section from 60° N to Charlie-Gibbs is the one section of the MAR where seamounts are formed without a strong hotspot signal in the geochemistry.

The line of seamounts along the Bight fracture zone (shown in Fig. 1 as an east-west line of seamounts at 57° N) is responsible for the peak in the 0–10 Myr and 10–20 Myr windows in Fig. 2. It seems likely that the fracture zone provided easy conduits for these volcanoes^{2,7}, although it is obvious that fracture zones are not necessary for seamount formation in the North Atlantic. There must be local perturbations to the plumbing system or other zones of weakness in the crust that control the location of the majority of the seamounts.

The first-order correlation of RCG seamounts with the geochemical data indicates that an injection of plume material is necessary for the formation of RCG seamounts. This suggests that in a normal spreading situation the MAR plumbing system has only enough pressure to lift magma to the top of the crust, which follows if decompression melting is the only source of magma. An injection of plume material overpressurizes the plumbing system; some of the excess pressure is released by the formation of seamounts, and some may be released by the formation of thick crust. The seamounts act like piezometers in the plumbing system and allow an estimation of the minimum overpressure resulting from the plume contribution. Seamounts with heights of ~1,500 fathoms are associated with the Charlie-Gibbs fracture zone, requiring an overpressure of ~42 MPa. For comparison, this is about half of the pressure in the Kilauea summit reservoir before the 1960 eruption on Kilauea's lower east rift zone³⁵. The North Atlantic is thought not to have a steady-state magma chamber, and Schilling²⁹ suggests that small, discontinuous, transient magma chambers may be more likely in slow-spreading ridges. If so, the overpressure almost certainly varies with distance along the ridge and with time, although it is not clear how these variations are reflected in the distribution of seamounts. The formation of seamounts between 60° N and the Charlie-Gibbs could be explained if the Iceland plume contribution pressurizes the MAR plumbing system as far south as the Charlie-Gibbs, but the plume material flows only as far as 60° N. A similar phenomenon is observed in Kilauea, where pressure changes in the summit reservoir are propagated down

the east rift zone before magma moves along the rift zone.

The model presented above suggests that the processes responsible for the generation of seamounts on the MAR are relatively simple, although the detailed variations in plume contributions and seamount distribution suggest that factors such as the volume of plume material, conduit geometry, location and offset of fracture zones, and the timing of plate reorganizations and hotspot interactions must also be considered. The North Atlantic may constitute the slow-spreading endmember of these processes, and it is not clear how, or if, these processes are applicable to faster spreading ridges.

This work was supported by the Office of Naval Research and the US Naval Oceanographic Office. During part of this work D.E. was at the Hawaii Institute of Geophysics, Honolulu. We thank Jean-Guy Schilling and Peter Vogt for helpful reviews of the manuscript.

Received 30 June; accepted 2 November 1988.

1. Batiza, R. *Geophys. Res. Lett.* **8**, 853-856 (1981).
2. Batiza, R. *Earth planet. Sci. Lett.* **160**, 195-206 (1982).
3. Jordan, T. H., Menard, H. W. & Smith, D. K., *J. geophys. Res.* **88**, 10508-10518 (1983).
4. Smith, D. K., thesis, Univ. of California (1985).
5. Smith, D. K. & Jordan, T. H. *Geophys. Res. Lett.* **14**, 1119-1122 (1987).
6. Smith, D. K. & Jordan, T. H. *J. geophys. Res.* **93**, 2899-2918 (1988).
7. Fornari, D. J., Batiza, R. & Luckman, M. A. in *Seamounts, Islands and Atolls* (Geophys. Monogr. 43 (eds. Keating, B. H. et al.) (Am. geophys. Un. Washington DC, 1987).
8. Abers, G. A., Parsons, B. & Weissel, J. K. *Earth planet. Sci. Lett.* **87**, 137-151 (1988).

9. Glenn, M. F. *Int. hydrogr. Rev.* **47**, 35-40 (1970).
10. Smoot, N. C. *Fall Tech. Mtg Am. Congr. of Surv. and Mapp.*, Niagara Falls, New York (1980).
11. Smoot, N. C. *Mar. Geol.* **47**, 153-163 (1982).
12. Smoot, N. C. *Mar. Geol.* **64**, 203-215 (1985).
13. Vogt, P. R. & Smoot, N. C. *J. geophys. Res.* **89**, 11085-11107 (1984).
14. Vogt, P. R. & Tucholke, B. E. in *The Geology of North America: M, The Western North Atlantic Region* 19-44 (Geol. Soc. Am., 1986).
15. Johnson, G. L. & Jakobsson, S. P. *J. geophys. Res.* **90**, 10073-10083 (1985).
16. Tucholke, B. E. in *The Geology of North America: M, The Western North Atlantic Region* 331-340 (Geol. Soc. Am., 1986).
17. Morgan, W. J. in *The Sea* **7**, 443-487 (Wiley, 1981).
18. Morgan, W. J. *Tectonophysics* **94**, 123-139 (1983).
19. Duncan, R. A. *J. geophys. Res.* **89**, 9980-9990 (1984).
20. Vogt, P. R. *Earth planet. Sci. Lett.* **13**, 153-160 (1971).
21. Vogt, P. R. in *Geodynamics of Iceland and the North Atlantic Area* (ed. Kristjansson, L.) 105-126 (Reidel, Dordrecht, 1974).
22. Vogt, P. R. *Earth planet. Sci. Lett.* **29**, 309-325 (1976).
23. Vogt, P. R. & Johnson, G. L. *Nature* **245**, 373-375 (1973).
24. Vogt, P. R. & Johnson, G. L. *J. geophys. Res.* **80**, 1399-1428 (1975).
25. Schilling, J.-G. *Nature* **242**, 565-571 (1973).
26. Schilling, J.-G. & Noe-Nygaard, A. *Earth planet. Sci. Lett.* **24**, 1-14 (1974).
27. Morgan, W. J. *J. geophys. Res.* **83**, 5355-5360 (1978).
28. Schilling, J.-G. *Nature* **314**, 62-67 (1985).
29. Schilling, J.-G. in *The Geology of North America: M, The Western North Atlantic Region* 137-156 (Geol. Soc. Am., 1986).
30. Rideout, M. L. & Schilling, J.-G. *Initial Rep. DSDP* **82**, 483-493 (1985).
31. Bougault, H. et al. *Earth planet. Sci. Lett.* **88**, 27-36 (1988).
32. Srivastava, S. P. & Tapscott, C. R. in *The Geology of North America: M, The Western North Atlantic Region* 379-404 (Geol. Soc. Am., 1986).
33. Klitgord, K. D. & Schouten, H. in *The Geology of North America: M, The Western North Atlantic Region* 351-378 (Geol. Soc. Am., 1986).
34. Vogt, P. R. in *The Geology of North America: M, The Western North Atlantic Region* 189-204 (Geol. Soc. Am., 1986).
35. Decker, R. W., Okamura, A. T., Dvorak, J. J. & Epp, D. *Eos* **64**, 901 (1983).

Geochemical evidence for invasion of Kilauea's plumbing system by Mauna Loa magma

J. M. Rhodes*, K. P. Wenz*, C. A. Neal†, J. W. Sparks* & J. P. Lockwood†

* Department of Geology/Geography, University of Massachusetts, Amherst, Massachusetts 01003, USA

† US Geological Survey, Hawaii National Park, Hawaii 96718, USA

From the beginning of the study of Hawaiian volcanism there has been controversy over possible relationships between the neighbouring active volcanoes Mauna Loa and Kilauea¹⁻⁵. Seismic activity, thought to reflect upward migration of magma, reveals that the magmatic plumbing systems apparently converge at depth to form a broad funnel within the mantle⁶. Although on rare occasions they have erupted concurrently, the brief historical eruptive record appears to show that when Kilauea is most active, Mauna Loa is in repose and vice versa, suggesting that they may be competing for the same magma supply^{5,7}. Petrological, geochemical and isotope data, however, require a diametrically opposite conclusion. Distinct differences in major-element, trace-element and isotope compositions of lavas are regarded as compelling evidence that the two volcanoes have separate magmatic plumbing systems, supplied by parental magmas from physically and geochemically distinct mantle sources⁸⁻¹³. Here we present preliminary geochemical data which show that in the past 2,000 years Kilauea has erupted a spectrum of lava compositions resembling historical Kilauea lavas at one end and Mauna Loa lavas at the other. We discuss the cause of this diversity, and speculate that magma from Mauna Loa may have invaded Kilauea's 'high-level' magmatic plumbing system.

Most surface lava flows on Kilauea are younger than 1 kyr (refs 14-16). To examine more of the recent prehistoric record, we sampled successive flows exposed in the walls of two pit craters, Kilauea Iki and Pauahi¹⁷. Kilauea Iki is close to the summit caldera; Pauahi is about 6 km down Kilauea's upper east rift zone. Age control for these flows depends in part on palaeomagnetic results^{14,15}, and in part on the age of the Uwekahuna Ash, a thin phreatomagmatic deposit derived from

Kilauea's summit^{4,15,18,19}. The Uwekahuna Ash is well exposed in the walls of Pauahi crater, about 45 m above the floor, but is not present in the walls of Kilauea Iki. From ¹⁴C measurements, Lockwood (personal communication) infers an age of 2.1 kyr for the ash. Using this age and palaeomagnetic results^{14,15}, we infer that lavas exposed in the walls of Pauahi crater range in age from greater than 2.1 to ~0.4 kyr, and those in Kilauea Iki range from 1.5-2.0 kyr to ~0.4 kyr.

Compositional differences between historical Kilauea and Mauna Loa lavas, which have been thoroughly documented⁸⁻¹³, are readily apparent from the ratios of Zr/Nb, Zr/Y and La/Yb (Figs 1 and 2), and from the averaged data summarized in Table 1. Previous studies propose that, at a given MgO content, prehistoric Kilauea lavas are more 'Mauna Loa-like' than their historical counterparts^{9,20}. Disputing this conclusion, Easton and Garcia²¹ found that Hilina Formation lavas older than 22 kyr exposed on Kilauea's south flank are compositionally close to modern lavas. Our data show that Kilauea lavas erupted over the past 2.0 kyr vary significantly in composition, particularly in abundances and ratios of the incompatible elements. In Fig. 1, historical lavas (AD 1823-1987) are characterized by Zr/Nb ratios of ~10 ± 1, whereas prehistoric lavas are more varied and display a distinctly bimodal distribution with maxima at ~11 and 14. The lavas with Zr/Nb ratios of ~14 are similar in both major- and trace-element composition to Mauna Loa lavas (Figs 1 and 2; Table 1). Prehistoric Kilauea lavas, with lower Zr/Nb ratios, show a range of compositions from those broadly resembling modern Kilauea (Zr/Nb < 11) and Hilina Formation lavas, to others with compositions intermediate (Zr/Nb = 11-12.5) between those of historical Mauna Loa and Kilauea lavas (Table 1; Fig. 2).

In Pauahi crater, the Mauna Loa-like flows are restricted to a few thin flow units immediately above the Uwekahuna Ash, and close to the top of the section (Fig. 2). The remainder of the section comprises flows with compositions closer to, but not identical with, modern Kilauea lavas¹⁷. Flows on the southern wall of Kilauea Iki crater grade from Kilauea-like compositions near the base, through intermediate compositions, to Mauna Loa-like flows in the middle of the section, and back through intermediate compositions to Kilauea-like flows at the top (Fig. 2)¹⁷.

To improve temporal control, we have also analysed samples of nine eruptive units from Kilauea's summit area, with ages

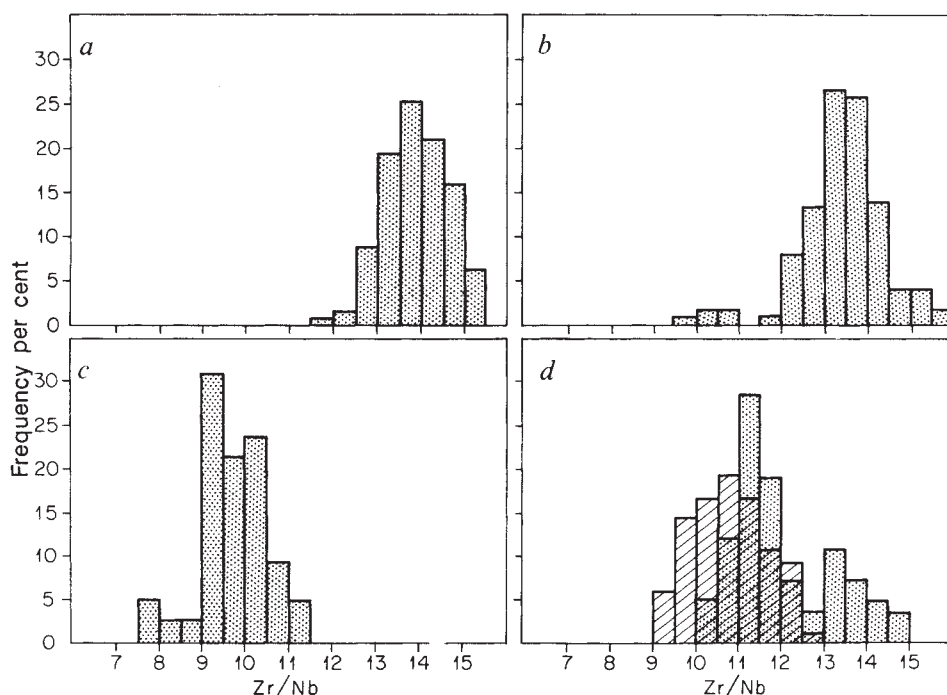


Fig. 1 Histograms of Zr/Nb ratios for historical (AD 1843–1984; $N = 137$) (a) and prehistoric (<30 kyr; $N = 130$) (b) Mauna Loa lavas, and historical (AD 1823–1987; $N = 42$) (c) and prehistoric (<2 kyr) as compared with historical Kilauea lavas. Note the greater range and bimodal distribution for recent prehistoric (<2 kyr) as compared with historical Kilauea lavas. Hilina Formation lavas (≥ 22 kyr) (hatched bars in d; $N = 41$) have Zr/Nb ratios that are close to those of historical lavas²¹.

Table 1 Comparison of average MgO-normalized major- and incompatible-trace-element abundances in prehistoric Kilauea lavas with their abundances in historical Kilauea and Mauna Loa lavas

Zr/Nb	Historical Mauna Loa lavas (>12.5)		Mauna Loa-like (>12.5)		Prehistoric Kilauea lavas Intermediate (11.0–12.5)		Kilauea-like (<11.0)		Historical Kilauea lavas (<11.0)	
	122/28		20/7		46/10		11/3		30/9	
Number of samples*	X	s.d.	X	s.d.	X	s.d.	X	s.d.	X	s.d.
SiO ₂	51.61	0.23	50.83	0.29	50.54	0.54	50.48	0.31	50.04	0.42
TiO ₂	2.08	0.06	2.27	0.06	2.49	0.17	2.51	0.07	2.65	0.13
Al ₂ O ₃	13.62	0.10	13.78	0.13	13.45	0.34	13.34	0.19	13.42	0.21
Fe ₂ O ₃ †	12.05	0.14	11.88	0.18	12.16	0.64	12.28	0.47	12.26	0.23
MnO	0.17	0.01	0.17	0.00	0.17	0.01	0.17	0.01	0.17	0.01
MgO	7.00	—	7.00	—	7.00	—	7.00	—	7.00	—
CaO	10.49	0.13	10.92	0.12	11.06	0.45	11.16	0.36	11.24	0.32
Na ₂ O	2.20	0.13	2.42	0.28	2.55	0.42	2.42	0.33	2.32	0.09
K ₂ O	0.39	0.04	0.40	0.02	0.44	0.05	0.46	0.02	0.51	0.04
P ₂ O ₅	0.24	0.02	0.23	0.01	0.24	0.02	0.25	0.02	0.27	0.02
Total	99.84	—	99.89	—	100.11	—	100.06	—	99.88	—
Rb	6.0	0.7	6.0	0.4	7.2	0.9	7.6	0.9	9.4	2.1
Sr	313	21.3	330	9.1	337	16.0	349	11.9	372	29.0
Nb	9.8	0.7	10.5	0.6	13.3	1.0	14.7	0.8	17.8	1.9
Zr	134	7.0	142	4.4	154	9.1	157	7.7	173	10.5
Hf	3.3	0.13	3.5	0.1	3.8	0.2	—	—	4.8	0.5
La	9.3	0.9	10.2	0.9	11.1	0.6	12.7	0.3	15.9	1.7
Ce	25.1	1.9	27.3	2.6	29.4	1.7	34.7	1.7	39.4	3.6
Sm	4.8	0.2	5.2	0.5	5.5	0.4	5.9	0.3	6.3	0.3
Eu	1.70	0.05	1.82	0.08	1.84	0.06	1.92	0.14	2.08	0.16
Yb	2.05	0.06	2.03	0.13	1.99	0.07	1.99	0.17	2.11	0.17
Lu	0.28	0.01	0.29	0.02	0.29	0.02	0.29	0.02	0.30	0.03
Y	23.6	0.8	24.5	0.7	24.9	1.1	25.4	1.2	25.2	1.0
Zr/Nb	13.7	0.6	13.6	0.6	11.6	0.4	10.7	0.2	9.8	0.6
Zr/Y	5.7	0.2	5.8	0.1	6.2	0.2	6.2	0.3	6.7	1.3
K/Y	136	12.4	136	8.0	147	16.5	151	10.2	168	11.6
Sr/Nb	32.0	1.4	31.6	1.8	25.5	2.0	23.8	1.0	20.9	2.7
(La/Yb) _{ch} ‡	2.7	0.3	3.0	0.2	3.4	0.3	3.9	0.4	4.6	0.4

The data have been normalized to a constant 7.0% MgO by addition or subtraction of olivine (Fo₈₈). Lavas that depart from an olivine-controlled trend have been excluded from the compilation. Major- and trace-elements Rb, Sr, Nb, Zr and Y were determined by X-ray fluorescence (XRF) at the University of Massachusetts. Rare-earth elements and Hf were determined by instrumental neutron activation analysis (INAA) in the laboratory of F. A. Frey at Massachusetts Institute of Technology. Data are from this study and refs 13, 17, 22 and 23. s.d., Standard deviation.

* Numbers refer to those analysed by XRF and INAA respectively.

† Total iron expressed as Fe₂O₃.

‡ Chondrite-normalized.

constrained by ^{14}C dating to between ~ 0.25 and 1.1 kyr. With the exception of the youngest flows (< 0.4 kyr), which resemble modern Kilauea lavas in composition, all these flows show compositional variations similar to the pit crater flows. Figure 2 and Table 1 show very clearly that there is a continuum of compositions for prehistoric Kilauea lavas ranging from those that resemble modern Kilauea lavas to those that resemble historical Mauna Loa flows. More have intermediate compositions than have either Kilauea-like or Mauna Loa-like compositions. Figure 3 illustrates a possible evolutionary scenario for the distribution of lavas of varied composition in Kilauea's summit and upper east rift zone.

Most Mauna Loa lavas are restricted in composition, clustering at the intersection of an olivine-controlled trend with a trend for more differentiated lavas. Typically, they contain ~ 6.8 – 8.0% MgO (refs 22, 23). Rhodes²³ attributes this uniformity to eruption from a continuously replenished shallow magma reservoir. The Mauna Loa-like lavas on Kilauea also share these characteristics, with MgO content between 6.6 and 8.6%. In contrast, Kilauea-like and intermediate lavas are more variable, ranging from differentiated basalts (5.5% MgO) to picrites (21% MgO)¹⁷.

The presence on Kilauea of flows with compositions that strongly resemble magma processed through Mauna Loa's shallow plumbing system might reflect flooding by Mauna Loa flows of the adjacent, and lower, Kilauea shield¹⁵. Such flows would have travelled over 20 km, originating on Mauna Loa's north-east rift zone. Far-travelled flows are typically dense tube-fed pahoehoe or thick aa units; the map relations, flow thickness, texture and vesicularity of these flows, on the other hand, indicate that they did not travel far, but were erupted on Kilauea. Furthermore, in our extensive investigations of Mauna Loa, no flows with intermediate compositional characteristics have been found throughout over 30 kyr of eruptive history^{9,23}.

This raises several intriguing possibilities. The magmas supplying Kilauea over the past 2 kyr may have been derived directly from the mantle either by partial melting of a mixed, multi-component mantle source, or by mixing of magmas from two or more distinct sources. Alternatively, they may have been supplied indirectly, as would be the case had magma from Mauna Loa invaded Kilauea's magmatic plumbing system.

The heterogeneous nature of the mantle source for Hawaiian volcanism is well established^{11,13,24,25}. However, it is mainly during the evolutionary stages of a volcano's history⁵, typically over time periods of a million years or more, that major temporal changes in lava composition occur. With the exceptions of Koolau²⁶ and Kahoolawe²⁷, the tholeiitic shield-building stages⁵ of most individual Hawaiian volcanoes are characterized by relatively uniform and distinctive geochemical features²⁸. Our unpublished data for Mauna Loa show that prehistoric lava compositions have remained comparatively homogeneous for over 30 kyr (Fig. 1). Similarly, on Kilauea, lavas of the Hilina Formation (> 22.5 kyr) remained relatively uniform in composition over perhaps as long as 50 kyr (ref. 21). We suggest, therefore, that compositional fluctuations in the mantle source within a time frame of less than 2.0 kyr are unlikely.

Although partial melting of a heterogeneous mantle source cannot be ruled out, we are persuaded otherwise by two key observations. First, Mauna Loa-like lavas erupted on Kilauea have all the compositional attributes of lavas processed through Mauna Loa's plumbing system. Second, these and intermediate-composition lavas first appear stratigraphically above the Uwekahuna Ash. The Uwekahuna Ash is thought to have been produced by a major explosive eruption associated with caldera collapse, which was itself initiated by massive magma drainage along one of the two rift zones^{4,15,18,19}. Evidence for extensive magma movement and drainage along Kilauea's east rift zone before the explosive 1924 summit eruption has been well documented^{4,5}. The possible magnitude of this process is illustrated by the recent discovery of voluminous (2-km^3), relatively young

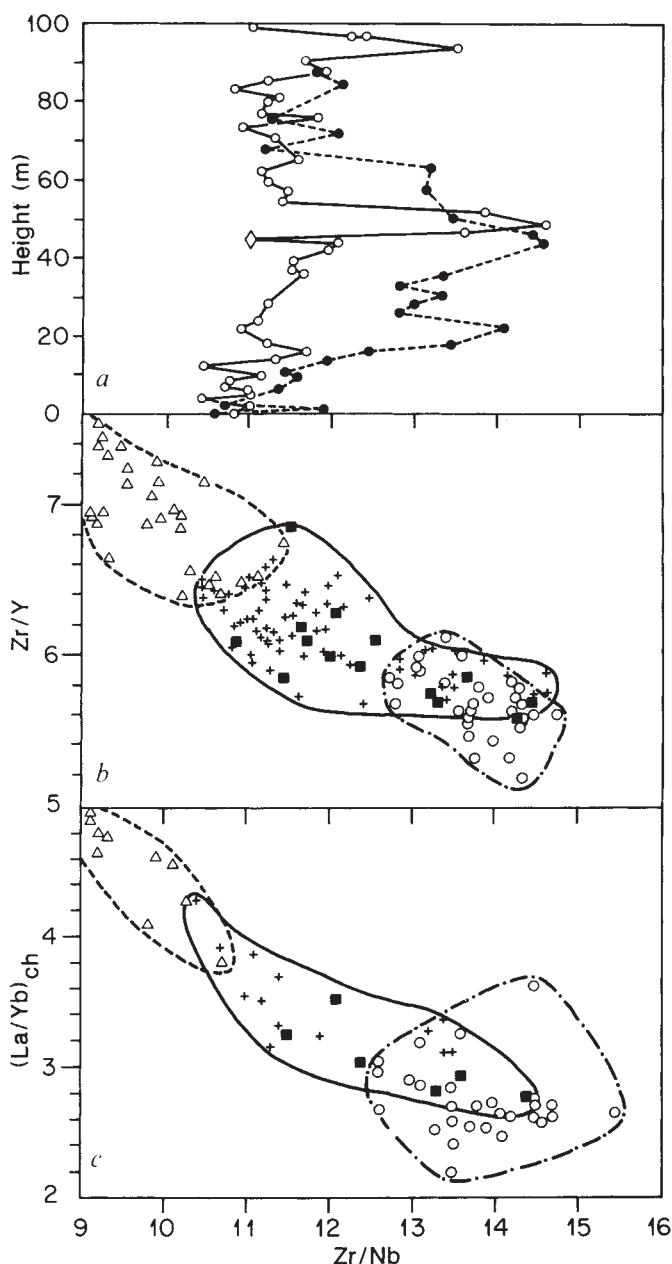


Fig. 2 *a*, Variation in Zr/Nb ratio with stratigraphic height in the walls of Pauahi (open circles joined by solid line) and Kilauea Iki (filled circles joined by dashed line) pit craters. Note the presence of the Uwekahuna Ash (open diamonds) at 45 m in the Pauahi section; this ash is presumed to be below the pit crater floor in the Kilauea Iki section. *b*, Zr/Y versus Zr/Nb ratios for Kilauea Iki and Mauna Loa lavas. Field for prehistoric Kilauea lavas shown as a solid line; ■ Prehistoric Kilauea lavas (200–1,100 yr BP) with ^{14}C age control; + Pauahi and Kilauea Iki lavas; △, Historical Kilauea lavas, field shown as dashed line; ○ Historical Mauna Loa lavas, field shown as dot-dashed line. *c*, Chondrite-normalized La/Yb versus Zr/Nb ratios for Kilauea and Mauna Loa lavas. Key is the same as for *b*. In *b* and *c* prehistoric Kilauea lavas reveal a range in composition from historical Kilauea to historical Mauna Loa.

lava flows emanating from the submarine extension of Kilauea's east rift zone, which are thought to represent such magma drainage²⁹.

We speculate that, following the voluminous magma drainage responsible for the Uwekahuna event, magma stored in the adjacent shallow plumbing system of Mauna Loa may have invaded Kilauea's evacuated plumbing system. Some of this

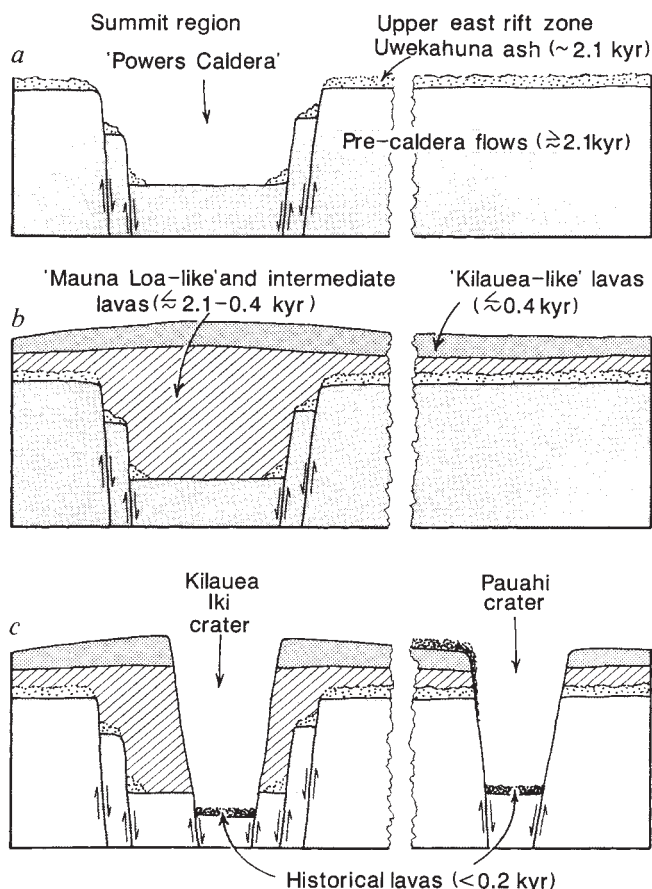


Fig. 3 Schematic representation of the inferred evolution of the geochemical variations in Kilauea summit and upper-east-rift-zone lavas over the past 2.1 kyr following the eruption of the Uwekahuna Ash. *a-c* represent successive configurations in time. The Uwekahuna Ash is a widespread phreatomagmatic base-surge deposit, derived from Kilauea's summit. It is thought to be associated with the collapse and formation of the ancestral 'Powers Caldera'^{4,15,18,19}. Lavas erupted before the formation of the Uwekahuna Ash and the Powers Caldera resemble modern Kilauea lavas in composition. Subsequent eruptions produced lavas that are more Mauna Loa-like in composition, or have compositions intermediate between those of modern Kilauea and Mauna Loa lavas. The greater thickness of these lavas in Kilauea Iki pit crater, in contrast with their restricted occurrence in Pauahi pit crater, is attributed to filling of the Powers Caldera and to the greater eruptive frequency in the summit region of the volcano compared with the upper east rift zone during this period. Only over the past 0.4 kyr have lavas resembling historical Kilauea lavas been erupted.

magma would retain its compositional identity in discrete magma batches and would subsequently be erupted. Much of the invading magma, however, would mix with residual pockets of Kilauea magma distributed throughout the plumbing system to produce magma with intermediate compositions. Table 1 and Fig. 2 show that such a mixing scenario is plausible. Subsequently, as Kilauea's magma supply recovered, eruptions of strictly Mauna Loa-like lavas would become increasingly rare and lavas of intermediate composition would predominate. Once initiated, such a magmatic connection might remain active for some time.

This model is at present based primarily on geochemical arguments. Seismic or geodetic observations show no evidence for currently active magma conduits between the two volcanoes, but there is tentative geophysical evidence for a fossil shallow magmatic connection. Aeromagnetic data display a roughly linear zone of high magnetic intensity from the vicinity of Kilauea caldera to the upper north-east rift zone of Mauna Loa,

and have been interpreted as indicating an interconnection between the two volcanoes^{30,31}. Furthermore, a recent seismic tomography study by Ho-Liu and Johnson (personal communication) identifies a sub-horizontal zone of high attenuation between the two volcanoes at a depth of 4–10 km that is roughly coincident with the aeromagnetic anomaly. As it is not presently active seismically, it clearly does not reflect a current magma conduit, but it could be the 'fossil' relict of the magma conduit required by our model.

We thank D. A. Clague and H. B. West for their thorough reviews. This work was supported in part by a faculty research grant from the University of Massachusetts, and in part by the NSF.

Received 3 June; accepted 24 November 1988.

1. Dana, J. D. *Characteristics of Volcanoes* 399 (Dodd, Mead Co., New York, 1890).
2. Dutton, C. E. *US geol. Surv. 4th A. Rep.* 75–219 (1884).
3. Daly, R. A. *Proc. natn. Acad. Sci. U.S.A.* 47, 47–122 (1911).
4. Jagger, T. A. *Geol. Soc. Am. Mem.* 21, 508 (1947).
5. MacDonald, G. A. & Abbott, A. T. *Volcanoes in the Sea* 441 (University of Hawaii Press, Honolulu, 1970).
6. Klein, F. W. *J. geophys. Res.* 87, 7719–7726 (1982).
7. Klein, F. W. *J. Volcan. geotherm. Res.* 12, 1–35 (1982).
8. Powers, H. A. *Geochim. cosmochim. Acta* 7 77–107 (1955).
9. Wright, T. L. *US geol. Surv. Prof. Pap.* 735, 39 (1971).
10. O'Nions, R. K., Hamilton, P. J. & Evenson, N. M. *Earth planet. Sci. Lett.* 34, 13–22 (1977).
11. Tatsumoto, M. *Earth planet. Sci. Lett.* 38, 63–87 (1978).
12. Leeman, W. P. *et al. Am. J. Sci.* 280A, 794–819 (1980).
13. Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* 1286 (Pergamon, Elmsford, New York, 1981).
14. Holcomb, R. T., Champion, D. E. & McWilliams, M. O. *Bull. geol. Soc. Am.* 97, 829–839 (1986).
15. Holcomb, R. T. *US geol. Surv. Prof. Pap.* 1350, 261–350 (1987).
16. Rubin, M., Lockwood, J. P. & Friedman, I. *US geol. Surv. Prof. Pap.* 1350, 209–243 (1987).
17. Wenz, K. P. thesis, Univ. Massachusetts (1988).
18. Powers, H. A. *Pacif. Sci.* 2, 278–292 (1948).
19. Decker, R. W. & Christiansen, R. L. *Explosive Volcanism: Inception, Evolution and Hazards* 122–132 (National Academy, Washington, DC, 1984).
20. Tilling, R. I., Wright, T. L. & Millard, H. T. Jr *US geol. Surv. Prof. Pap.* 1350, 641–689 (1987).
21. Easton, R. M. & Garcia, M. O. *Bull. volcan.* 43, 657–673 (1980).
22. Rhodes, J. M. *J. geophys. Res.* 88A, 869–879 (1983).
23. Rhodes, J. M. *J. geophys. Res.* 93, 4453–4466 (1988).
24. Chen, C.-Y. & Frey, F. A. *Nature* 302, 785–789 (1983).
25. Stille, P., Unruh, D. M. & Tatsumoto, M. *Nature* 304, 25–29 (1983).
26. Roden, M. F., Frey, F. A. & Clague, D. A. *Earth planet. Sci. Lett.* 69, 141–158 (1984).
27. West, H. B. *et al. Nature* 330, 216–220 (1987).
28. Frey, F. A. & Roden, M. F. *Mantle Metasomatism* (eds Menzies, M. A. & Hawkesworth, C. J.) 423–463 (Academic, London, 1987).
29. Holcomb, R. T. *et al. Geology* 16, 400–404 (1988).
30. Malahoff, A. & Woolard, G. P. *Pacif. Sci.* 20, 265–311 (1966).
31. Flanagan, V. J. & Long, C. L. *US geol. Surv. Prof. Pap.* 1350, 935–946 (1987).

Fitness costs of gestation and lactation in wild mammals

T. H. Clutton-Brock, S. D. Albon & F. E. Guinness

Large Animal Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Like a number of plants^{1,2}, some mammals commonly produce more progeny than they can afford to rear, terminating investment in some or even all of their offspring once the resources available for breeding are known^{3–5}. Adaptive interpretations of juvenile wastage rely on the argument that the costs of gestation are small compared to those of feeding offspring. Though energetic evidence supports this conclusion⁶, it is unsafe to assume that the relative costs of gestation and lactation to the mother's survival and future reproductive success follow the same pattern because lactation commonly coincides with the period of maximum food availability. Controlling for individual variation, we show that in wild red deer (*Cervus elaphus* L) any costs of gestation to the mother's subsequent survival and reproductive success are slight compared to those of lactation.

Reproductive histories of 283 individually recognizable hinds were collected in a food-limited population on the Isle of Rhum, Inner Hebrides between 1971 and 1987 (ref. 7). In each year of

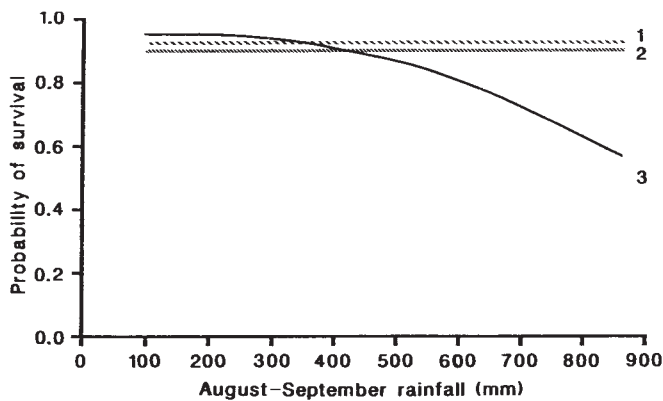


Fig. 1 Logistic curves illustrating the probability of a hind surviving a winter following a year in which she was (1) barren, (2) had produced a calf but lost it shortly after birth, and (3) reared a calf until weaning, are plotted against August–September rainfall (mm), which depresses survival in this population ($\chi^2=23.3$, $P<0.001$) (ref. 9). These curves are standardized for hind age and values shown are for 12-year-olds (quadratic in age: $\chi^2=78.1$, d.f.=2, $P<0.001$). N (hind-years)=1,479. Coefficients in the exponential equation (see text) are: (1) $3.107-0.000907$ (August–September rainfall); (2) $2.444-0.000177$ (August–September rainfall); (3) $3.563-0.003427$ (August–September rainfall).

their lives, hinds were allocated to four categories. The first category consisted of those hinds that produced no calf (barren). Previous evidence suggests that when a hind was barren she had usually failed to conceive⁷ and consequently did not suffer the costs of either gestation or lactation. The second category consisted of hinds that lost their calf within three months of its birth. In practice, 90% of the 156 such calves died within their first two weeks of life. In these years, mothers consequently avoided most of the energetic cost of lactation. The third category consisted of hinds whose calf had died between 4 and 12 months of age. 80% of the 204 calves allocated to this category died between February and May aged 9–12 months. Finally, the fourth category consisted of hinds that had successfully raised their calf to at least 12 months of age. We investigated relationships between the period for which the calf survived and four components of the mother's fitness: her survival during the winter following the calf's birth; her probability of producing a calf in the year after the calf's birth (fecundity); and the birth date and weight of her subsequent calf, both of which are known to affect calf survival⁸. In this population, high rainfall in August and September depresses the growth of grass and reduces food availability at the onset of winter, with the result that both fecundity and over-winter survival fall after wet autumns⁹. Our analysis consequently allowed for annual variation in rainfall in late summer.

Differences in fitness components between barren hinds and those whose calves died soon after birth should reflect the costs of gestation, whereas differences between mothers whose calves died soon after birth and hinds whose calves survived for six months or more should reflect the costs of lactation. Estimates of the costs of reproduction based on comparisons of this kind can underestimate the actual costs of breeding because, despite their greater energetic expenditure, mothers of superior phenotypic quality sometimes show higher survival rates or superior reproductive performance when compared to inferior phenotypes that did not breed in the previous year^{10,11}. To minimize this problem, we compared the subsequent reproductive performance of the same hinds following years when they had been barren, had produced calves but lost them soon after birth or had reared calves to over six months.

In these analyses we included a factor to remove the effects of individual differences between mothers and restricted our

sample to 145 females born between 1967 and 1978 to ensure that we included reasonable samples of breeding attempts for each individual. Generalized linear models¹² were used to estimate the probability of the mother's survival or fecundity in relation to her reproductive status in the previous year¹³. The dependent variables were binary (survive = 1, die = 0; fecund = 1, not fecund = 0) and must be related to any explanatory variable (such as summer rainfall) in a nonlinear way. A logistic model¹⁴ was used to relate a particular transformation of the probability of survival or breeding for each individual (= the logarithm of the odds against dying or failing to breed) to a linear combination of explanatory variables. Logistic regressions were of the form:

$$P(Y_i) = \frac{\exp(A + B_1x_{i1} + B_2x_{i2} + B_3x_{i3} + B_4x_{i4} + B_5x_{i5})}{1 + \exp(A + B_1x_{i1} + B_2x_{i2} + B_3x_{i3} + B_4x_{i4} + B_5x_{i5})}$$

where A , B_1 , B_2 , ..., B_5 = constants, x_{i1} = August–September rainfall (mm), x_{i2} = previous reproductive status, x_{i3} = reproductive status by rainfall interaction, x_{i4} = age (years), x_{i5} = age² and $i = 1, 2, 3, \dots, N$ hind-years. Offspring birth weights and log-transformed calving dates were examined using an analysis of variance.

The fitness costs of gestation were evidently slight compared to those of lactation. In all years, the survival of barren females did not differ from those whose calves died during their first summer ($\chi^2=0.403$, NS). Except in years when late summer rainfall was unusually low and food availability in autumn was unusually high, both these categories of hinds were more likely to survive than mothers whose calves survived until weaning ($\chi^2=17.002$, $P<0.001$, see Fig. 1; reproductive status by rainfall interaction: $\chi^2=6.046$, $P<0.05$).

The subsequent fecundity of hinds did not differ between years when they had been barren and years when they had produced calves but lost them shortly after birth ($\chi^2=2.045$, NS). In contrast, their chances of calving were reduced after years in which they had suffered the costs of lactation (barren and early calf loss versus calf reared to >6 months, $\chi^2=292.21$, $P<0.001$, see Fig. 2) and this difference increased with rainfall

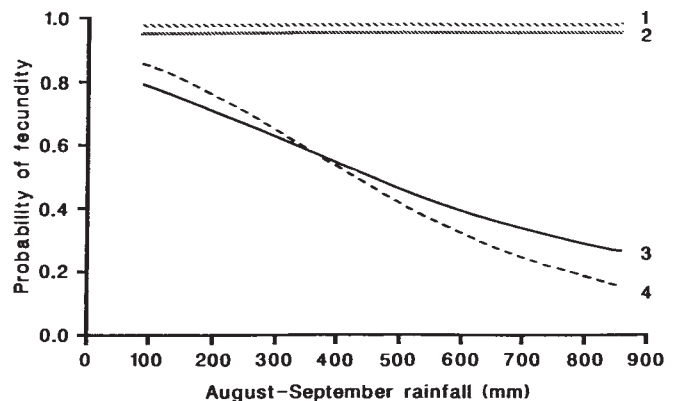


Fig. 2 Logistic curves illustrating the probability of a hind producing a calf in the summer following years when she was (1) barren, (2) had produced a calf but lost it shortly after birth, (3) had reared a calf for at least 6 months but lost it in late winter or spring, and (4) had reared a calf successfully, are plotted on August–September rainfall, which affects fecundity in this population ($\chi^2=29.1$, $P<0.001$) (ref. 9). These curves are standardized for three variables that affect fecundity in this population: winter temperature ($\chi^2=12.2$, $P<0.001$), values shown are for mean December–February daily temperature (4.5 °C); hind age (quadratic in age, $\chi^2=32.0$, d.f.=2, $P<0.001$), values shown are for 12-year-old hinds; and individual identity ($\chi^2=298.3$, d.f.=134, $P<0.001$), set here to the median individual. Coefficients in the exponential equation (see text) are: (1) $3.728-0.000285$ (August–September rainfall); (2) $3.319-0.000055$ (August–September rainfall); (3) $1.637-0.003205$ (August–September rainfall); (4) $2.160-0.004745$ (August–September rainfall).

Table 1 Mean calving dates and offspring birth weights of red deer hinds

	Subsequent calving date	Birth weight of next calf (kg)
Barren	5 June	6.614
Calf produced but dies soon after birth	3 June	7.029
Lactating:		
Calf reared for at least 6 months but died between 6 and 12 months	14 June	6.300
Calf survives to 1 year old	14 June	6.466

Mean calving dates and offspring birth weights of hinds following years in which they had been barren, produced a calf but lost it shortly after birth, reared a calf for at least 6 months but lost it between 6 and 12 months, and reared a calf which survived to one year. Means were standardized to the average individual and average year (1979 for calving date and 1978 for birth weight). In the case of offspring birth weight, means were adjusted for differences in calf sex ratio and the table shows values for males, which are born 0.38 kg heavier than females. NS, not significant; *, $P < 0.05$; *** $P < 0.001$.

in late summer (reproductive status by rainfall interaction, $\chi^2 = 13.1$, $P = 0.001$). Similarly, the subsequent calving date of hinds did not differ between years when they had been barren and years when they had produced calves but lost them shortly after birth ($t = 1.14$, NS), whereas on average they calved and, presumably, conceived ten days later when their previous calf survived to weaning ($t_{436} = 7.61$, $P < 0.001$; see Table 1). This change probably represented an appreciable cost, for a calf's chances of surviving the first year of life declined by 1% per day¹⁵.

Finally, hinds produced heavier calves after years when they had been barren or lost their calves shortly after birth than following years when they had reared calves into the winter ($F_{2,286} = 5.517$, $P < 0.001$; see Table 1). However, hinds produced heavier calves following years when they had produced but lost calves than following those when they had been barren ($t_{102} = 2.04$, $P < 0.05$), presumably because failure to breed was associated with poor body condition and this had an effect on the hind's reproductive performance the following season.

Our results thus confirm that in food-limited populations the fitness costs of gestation are slight compared to those of lactation. It may often be to the advantage of females to produce more offspring than they can usually rear and subsequently to reduce the number of offspring to a level that they can support. This may explain why, in many species, juvenile survival declines more rapidly with resource shortage or with rising population density than fecundity¹⁶.

We thank the Director of the Nature Conservancy Council for permission to work on Rhum; the Council's staff on the island and in Inverness for their support; SERC, NERC and Royal Society for financial support; and David Brown, Andrew Laurie and Paul Harvey for comments or advice.

Received 27 July; accepted 9 December 1988.

- Stephenson, A. G. *Ecol. Rev.* **61**, 57–64 (1980).
- Stephenson, A. G. *A. Rev. Ecol. Syst.* **12**, 253–279 (1981).
- Rivers, J. P. W. & Crawford, M. A. *Nature* **252**, 297–298 (1974).
- Gosling, L. M. *Am. Nat.* **127**, 772–795 (1986).
- Tait, D. E. N. *Am. Nat.* **115**, 800–808 (1980).
- Ofstedahl, O. T. *Symp. Zool. Soc. Lond.* **51**, 33–85 (1984).
- Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: the Behaviour and Ecology of Two Sexes* (University of Chicago Press, Chicago, 1982).
- Guinness, F. E., Clutton-Brock, T. H. & Albon, S. D. *J. Anim. Ecol.* **47**, 817–832 (1978).
- Albon, S. D. & Clutton-Brock, T. H. in *Ecological Change in the Uplands* (eds Usher, M. B. & Thompson, D. B. A.) 93–107 (Blackwell, Oxford, 1988).
- Clutton-Brock, T. H. & Harvey, P. H. *Proc. R. Soc. B* **205**, 547–565 (1979).
- Partridge, L. & Harvey, P. H. *Nature* **316**, 20 (1985).
- McCullough, P. & Nelder, J. A. *Generalized Linear Models* (Chapman & Hall, London, 1983).
- Albon, S. D., Mitchell, B., Huby, B. J. & Brown, D. J. *Zool.* **209**, 447–460 (1986).
- Cox, D. R. *The Analysis of Binary Data* (Methuen, London, 1970).
- Clutton-Brock, T. H., Major, M., Albon, S. D. & Guinness, F. E. *J. Anim. Ecol.* **56**, 53–67 (1987).
- Fowler, C. H. *Current Mammalogy* **1**, 401–441 (1987).

The potential role of pathogens in biological control

Michael E. Hochberg

Department of Pure and Applied Biology,
Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK

It is now well established that pathogens such as viruses, fungi, bacteria and protozoans can have profound effects on the dynamics of their invertebrate host populations¹. Theoretical^{1,2} models of invertebrate host–pathogen interactions which assume uniform structure of the pathogen population may reasonably explain the oscillatory behaviour observed in some systems³, but do not adequately describe the existence of more constant populations found in other host–pathogen interactions^{4–7}. An examination of the literature relating to these relatively stable systems suggests that the common thread is the eventual transmission of some of the more protected, longer-lived stages of the pathogen occurring in reservoirs, such as the soil^{4,5}, host cadavers on trees⁸, or the live host itself⁷. In this letter, I propose a new theoretical model which incorporates this population structure and accounts for the range of dynamics observed in natural systems. In particular, I show that host populations may be regulated to low and relatively constant densities if sufficient numbers of pathogens are translocated from pathogen reservoirs to habitats where transmission can occur. An understanding of pathogen reservoirs may be of value in the design of biological control programmes and may greatly increase the effectiveness of pathogens as biological control agents.

Models of invertebrate host–pathogen interactions often assume that the transmissibility and lifespan of the pathogen does not differ between pathogen individuals, or stages^{1,2}. In nature, however, pathogen populations are far from uniform, because heterogeneities can exist in their spatial and genetic composition⁹. Models which assume homogeneity in pathogen populations have been used to explain the periodic cyclic behaviour of some forest Lepidoptera species^{2,3}, but the biological features of the pathogen responsible for these oscillatory dynamics (namely high pathogenicity and long-lived stages) are also characteristic of some systems in which the pathogen regulates its host to low and relatively constant densities^{5,8,10}.

The model presented here (Fig. 1) considers a more realistic structure of the pathogen population than earlier models by dividing the pathogen into two distinct subpopulations, one transmissible and short-lived, and the other non-transmissible and long-lived. For example, in the spruce sawfly, *Gilpinia hercyniae*, the transmissible stages of its nuclear polyhedrosis virus reside on the surface of spruce needles where they are exposed to vagaries of the environment, and are thus highly vulnerable to removal or to mortality^{8,11}. Alternatively, stages of the virus can reside in more protected and remote habitats, such as under the bark of the trees, in the upper layers of the soil or in the cadavers of virus-killed hosts⁸. Stages in these 'reservoirs' can only cause infections if they are somehow translocated to the surface of the spruce needles by abiotic and biotic agents such as wind, rain, birds or the live host itself^{8,12}. Although an obvious simplification of the real system, this scenario captures the essential dichotomy found in many invertebrate herbivore–pathogen interactions^{5,8,13–17}.

For the purposes of this study the invertebrate host is assumed to have continuous and overlapping generations, but the underlying results of the model also hold true for host taxa with discrete and non-overlapping generations. In addition, for the clarity of this presentation, the following assumptions are made¹: (1) the risk of infection to a given host increases linearly with the density of the transmissible pathogen; (2) the production of pathogen stages in the host individual is directly proportional to the average lifespan of the infected host; (3) the infection is invariably lethal; (4) secondary infections are rare or do not

occur at all; and (5) the direct transmission of protected stages to susceptible hosts does not occur. Relaxing these assumptions does not change the basic conclusions of this study.

The dynamics of the host-pathogen interaction are described in terms of four differential equations which capture, in a general way, population processes which are similar to those described above in the invertebrate herbivore-pathogen scenario. These equations are for the density of susceptible hosts (X), infected hosts (Y), transmissible pathogen stages (W) and non-transmissible stages (Q) (Fig. 1): $dX/dt = (a_0 - b)X + a_1Y - \beta XW$; $dY/dt = \beta XW - (\alpha + b - a_2)Y$; $dW/dt = \theta_1Y - (\mu + \lambda)W + \nu Q - \phi\beta XW$; and $dQ/dt = \theta_2Y - (\rho + \nu)Q + \lambda W$. Here a_0 and b are the per capita birth and death rates of susceptible hosts. Infected hosts give birth to either healthy or infected offspring at per capita rates of a_1 and a_2 , respectively. The transmission parameter is β , and when infection takes place an average of ϕ stages are removed from the transmissible subpopulation of the pathogen. Once infected, the host can succumb either to natural causes of mortality (b) or to disease-related death (α). As the number of pathogen stages produced per infected host is assumed to be directly related to the average lifespan expected of the infected host^{1,2} (that is, $1/(\alpha + b)$), the release of infective stages into the transmissible subpopulation and the reservoir can be represented by the steady states of θ_1Y and θ_2Y respectively. The transmissible stages die at a per capita rate μ or are translocated at a constant per capita rate of λ to the reservoir.

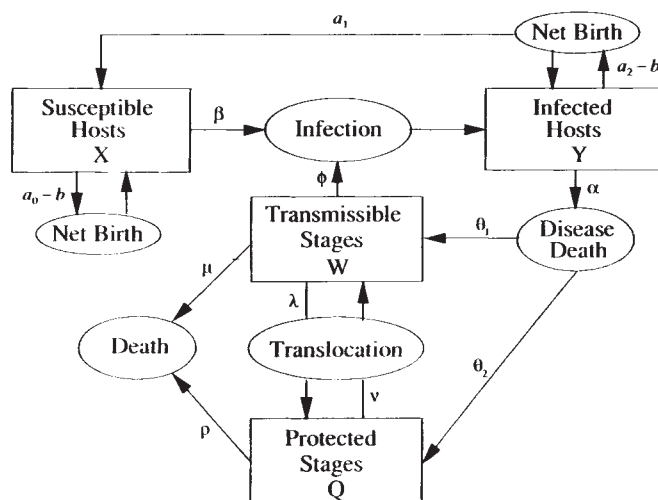


Fig. 1 Diagrammatic flow chart of transitions and rates in the interactions of susceptible hosts (X), infected hosts (Y), exposed, transmissible stages of the pathogen (W), and protected, non-transmissible stages (Q).

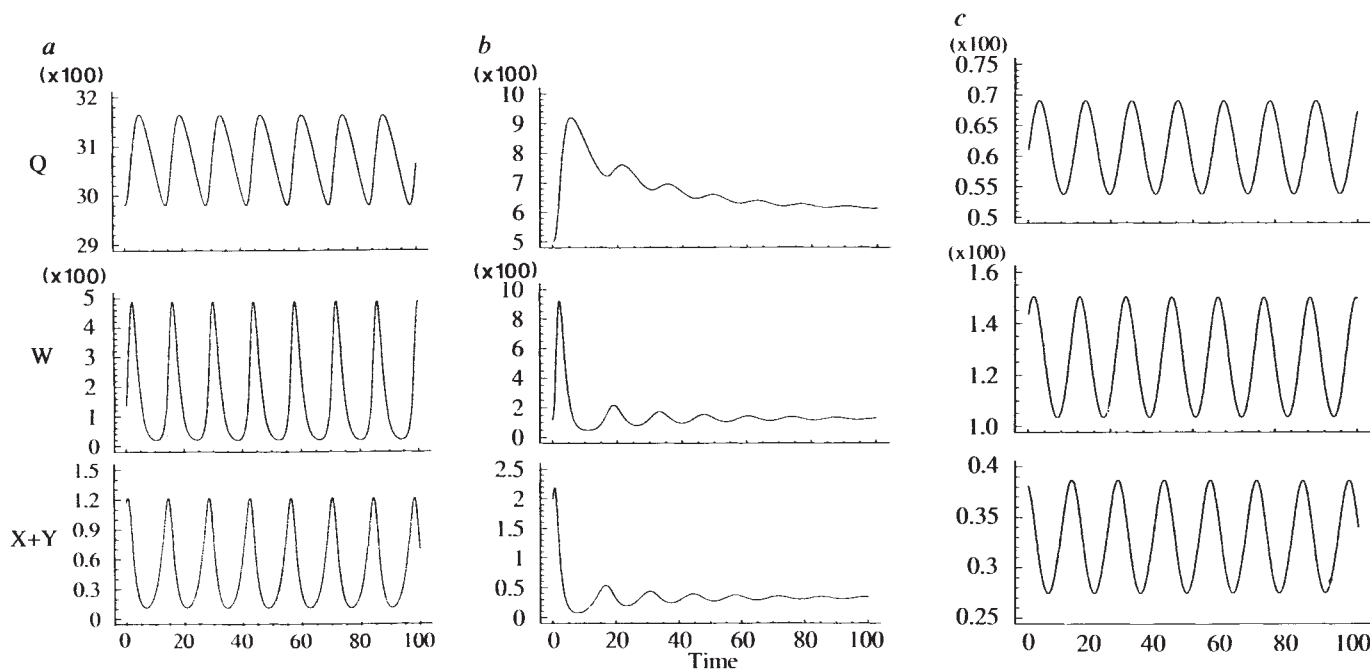


Fig. 2 Examples of population behaviour of total hosts ($X + Y$), transmissible pathogen stages (W) and protected stages (Q) for three different rates of flow of pathogen stages from the protected to the transmissible pathogen subpopulations. *a*, Low rates of flow ($\nu = 0$) resulting in oscillatory populations; *b*, intermediate rates of flow ($\nu = 0.04$) resulting in constant populations; *c*, high rates of flow ($\nu = 0.4$) giving rise to oscillatory populations. The effect of the pathogen reservoir on the host's dynamics can be intuitively understood as follows. First imagine a situation in which pathogen stages are not translocated from the reservoir to the transmissible subpopulation (corresponding to $\nu = 0$). The dynamics of the interaction are similar to those found by Anderson and May^{1,2}: persistent oscillations with a period determined to some extent by the rate of liberation of pathogen stages into the transmissible subpopulation (θ_1) and the average lifespan of transmissible stages ($1/(\mu + \lambda)$). Add to this an intermediate flow of stages from the reservoir to the transmissible subpopulation. Now, at peak populations of the transmissible pathogen, stages are being translocated to and stored in the more protected habitats. At times when the transmissible stages are sparse, the release effects of the reservoir result in a net flow of stages to the transmissible subpopulation. Thus, the storage and release capacities of the reservoir tend to 'buffer' the transmissible subpopulation of the pathogen through time, resulting in constant populations of both host and pathogen at equilibrium. If the flow from the protected to the transmissible subpopulation is too strong, then the two pathogen subpopulations effectively become a single entity and the host and pathogen interact in much the same way as they did in the first scenario (that is, oscillatory populations). The important difference is that the greater input of pathogen stages into the transmissible subpopulation has the effect of depressing the host population to lower levels than in the first case (compare Fig. 2*a* and *c*). Parameters common to all three simulations are $a_0 - b = 0.5$; $a_1 = 0.2$; $\beta = 0.005$; $\phi = 1$; $\alpha + b - a_2 = 1$; $\theta_1 = 10^3$; $\theta_2 = 0$; $\lambda = 0.2$; $\mu = 1$; and $\rho = 0.001$.

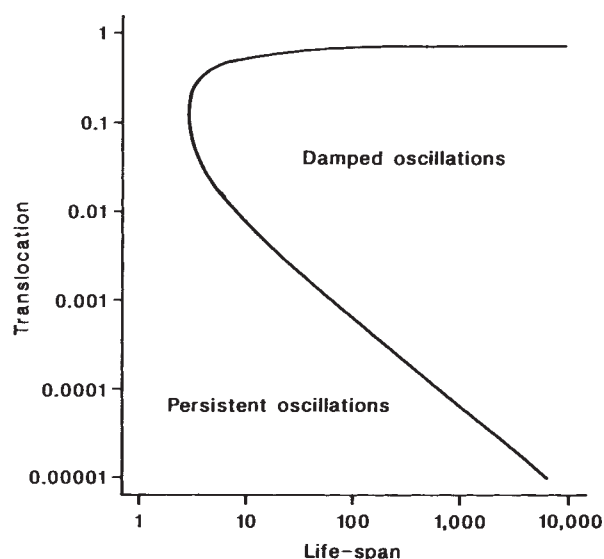


Fig. 3 Population dynamics as determined by the average lifespan of pathogen stages in the reservoir $1/\rho$, and the rate of translocation from the reservoir to transmissible subpopulation ν . Two types of population behaviours occur: oscillatory (unshaded) and constant (shaded) trajectories. Although the boundaries of oscillatory and constant populations will depend to some extent on the values of the other model parameters, there is a simple linear relationship between the logarithms of $1/\rho$ and ν through most of the parameter space. Estimation of these two crucial parameters^{25,26} can be made from measurements of the rate of translocation of the pathogen from the reservoir to the transmissible subpopulation (ν) and the average duration of the activity of the pathogen in the reservoir ($1/\rho$ or $1/(\rho + \nu)$). Anderson²⁷ has found that if the extent of artificial introduction of a pathogen into the host population is sufficient, then it can result in the dampening of cyclic populations. This is very similar to the effect of adequate translocation from the pathogen reservoir to the transmissible subpopulation. Values for other parameters are as given for Fig. 2, except that $\lambda = 1$.

In a similar way, stages residing in the reservoir die (ρ) or are translocated at a constant per capita rate ν to the transmissible subpopulation.

The properties of the model are straightforward. There is a single positive equilibrium only if the pathogen is sufficiently virulent (that is, $\alpha > a_1 + a_2 - b$) and if the number of stages removed from the transmissible subpopulation when infection occurs is sufficiently small. As other forms of population regulation are not included in the model, in the absence of the disease the host population grows unregulated at an exponential rate, $a_0 - b$.

As illustrated in Figs 2 and 3, the model exhibits one of two distinct types of behaviour at equilibrium: (1) stable oscillations or (2) constant population densities. The basic features of the model that determine which of these prevail are the storage and release of pathogen stages in the reservoir (governed primarily by constants λ , θ_2 , $1/\rho$ and ν). Very low values of these potentials give rise to stable oscillations (Figs 2a and 3). As they are increased to intermediate levels, the result is constant host and pathogen populations (Figs 2b and 3). Finally, increasing the release potential of the reservoir (governed by the constant ν) even further produces, once again, cyclic populations (Figs 2c and 3). But further increases in the storage of stages (governed by the constant $1/\rho$) continues to enlarge the parameter space for constant host populations (Fig. 3). Changes in other parameters of the model may affect the boundaries of the stable and unstable equilibrium, but do not alter the central results.

There are number of real examples which support the results presented here^{4-7,13-15}. The most striking of these comes from the interaction between Hepialid moths of the genus *Wiseana* and their nuclear polyhedrosis virus^{4,10,18-21}. The larvae of these

moths are pests of grasses in pastures in New Zealand. In long-term pasture leys, larvae that die from the disease accumulate in the upper layers of the soil to form a reservoir. Translocation of the virus from this reservoir to the larval food source (grass swards) has been attributed to the grazing of animal stock. The populations of the pest in these pastures are low and fairly constant from year to year. But management practices which disturb the pathogen reservoir (such as ploughing the pastures) and/or reduce the translocation of the pathogen from the reservoir to the grass swards (no grazing, for example) result in outbreaks of the pest, with highly variable populations from year to year. If both of these disruptive practices are halted, then the pathogen reservoir slowly builds, accompanied by the depression of the pest population. Thus, the creation of a distinct reservoir along with specific mechanisms for mobilization of the pathogen into the host population seems to mediate the differences in dynamical behaviour, which is in agreement with the model proposed here. Other real systems where pathogen reservoirs have an apparently similar role in the biological control of arthropod pests are: the Japanese beetle, *Popillia japonica*, and its bacterium, *Bacillus popilliae*⁵; the spruce sawfly, *Gilpinia hercyniae*, and its nuclear polyhedrosis virus⁶; the palm rhinoceros beetle, *Oryctes rhinoceros*, and its rhabdovirus⁷; the velvetbean caterpillar, *Anticarsia gemmatilis*, and its nuclear polyhedrosis virus¹³; the grasshopper, *Zonocerus variegatus*, and its fungus, *Entomophthora grylli*¹⁴; and the cabbage looper, *Trichoplusia ni*, and its nuclear polyhedrosis virus¹⁵. It should be stressed that, in these examples, the details of the mathematical model (for example, the amplitude and period of the cycles) are very difficult to corroborate because of the requirements of long runs of population data and accurate estimates of the parameters of the model. Rather, the crucial prediction made by the model is a strategic one: increasing the levels of release of pathogen stages from established reservoirs to the habitat of the host should lead to lower and more constant populations.

The model presented here constitutes a realistic framework for predicting the consequences of potential strategies in biological control programmes. Equally important, the model serves to clearly identify the biological factors that should be measured and manipulated, whether for experimentation or the actual practice of biological control. Measures²² to increase the efficacy of indigenous pathogens of arthropod pests should focus on the identification and manipulation of pathogen reservoirs and the processes that move pathogens between non-transmissible and transmissible subpopulations. In particular, the conservation of potential vectors of the pathogen should be paramount. For the introduction of exotic pathogens as 'classical' biological control agents, the model suggests that the conditions for determining their likelihood of success are (1) the long lifespan of pathogen stages residing in reservoirs and (2) the propensity of these stages to be translocated to the habitat of the host for possible transmission. Thus, the screening of promising agents and systems with pathogen reservoirs in mind is a sensible practice. In the past, microbial biological control programmes^{23,24} have focused on measures which yield short-term control of arthropod pests. These results indicate that long-term control is possible and may not entail much cost or effort.

I thank R. M. Anderson, P. F. Entwistle, H. J. C. Godfray, M. P. Hassell, R. M. May, C. Prior and J. K. Waage for helpful comments during the preparation of the manuscript. This work was supported by an ORS award from the Committee of Vice-Chancellors and Principals of the Universities of the UK.

Received 1 July; accepted 22 November 1988.

1. Anderson, R. M. & May, R. M. *Phil. Trans. R. Soc. B* **291**, 451-524 (1981).
2. Anderson, R. M. & May, R. M. *Science* **210**, 658-661 (1980).
3. Baltensweiler, W. *Can. Ent.* **96**, 792-800 (1964).
4. Crawford, A. M. & Kalmakoff, J. J. *Invertebr. Path.* **29**, 81-87 (1977).
5. Dutky, S. R. in *Insect Pathology: an Advanced Treatise* Vol. 2 (ed. Steinhaus, E. A.) (Academic, New York, 1963).
6. Neilson, M. M., Martineau, R. & Rose, A. H. *Comm. Inst. Bio. Control Tech. Commun.* **4**, 136-143 (1971).

7. Bedford, G. O. *A. Rev. Ent.* **25**, 309-339 (1980).
8. Evans, H. F. & Entwistle, P. F. in *Microbial and Viral Pesticides* (ed. Kurstak, E.) (Dekker, New York, 1982).
9. Tanada, Y. & Fuxa, J. R. in *Epizootiology of Insect Disease* (eds Fuxa, J. R. & Tanada, Y.) (Wiley, New York, 1987).
10. Moore, S. G., Kalmakoff, J. & Miles, J. A. R. *N.Z. J. Zool.* **1**, 85-95 (1973).
11. Bird, F. T. *J. Insect Path.* **3**, 352-380 (1961).
12. Entwistle, P. F., Adams, P. H. W. & Evans, H. F. *J. Invertebr. Path.* **29**, 354-360 (1977).
13. Young, S. Y. & Yearian, W. C. *Envir. Ent.* **15**, 573-580 (1986).
14. Chapman, R. F. & Page, W. W. *J. Anim. Ecol.* **48**, 271-288 (1979).
15. Jaques, R. P. *J. Invertebr. Path.* **23**, 140-152 (1974).
16. Ignoffo, C. M., Hostetter, D. L., Sikorowski, P. P., Sutter, G. & Brooks, W. M. *Envir. Ent.* **6**, 411-415 (1977).
17. Thompson, C. G. & Scott, D. W. *J. Invertebr. Path.* **33**, 57-65 (1979).
18. Moore, S. J. *N.Z. J. Agric.* **124**, 39-41 (1972).
19. Fleming, S. B., Archibald, R., Stewart, K. & Kalmakoff, J. *Proc. 3rd Australasian Conf. Grassland Invertebr. Ecol.* (ed. Lee, K. E.) (Government Printer, S. Australia, 1982).
20. Kalmakoff, J. & Crawford, A. M. in *Microbial and Viral Pesticides* (ed. Kurstak, E.) (Dekker, New York, 1982).
21. Fleming, S. B., Kalmakoff, J., Archibald, R. D. & Stewart, K. M. *J. Invertebr. Path.* **48**, 193-198 (1986).
22. Ignoffo, C. M. *J. Invertebr. Path.* **31**, 1-3 (1978).
23. Burges, H. D. *Microbial Control of Pests and Plant Diseases 1970-1980* (Academic, New York, 1981).
24. Kurstak, E. *Microbial and Viral Pesticides* (Dekker, New York, 1982).
25. Fuxa, J. R. in *Epizootiology of Insect Diseases* (eds Fuxa, J. R. & Tanada, Y.) (Wiley, New York, 1987).
26. Evans, H. F., Bishop, J. M. & Page, E. A. *J. Invertebr. Path.* **35**, 1-8 (1980).
27. Anderson, R. M. *Parasitology* **84**, 3-33 (1982).

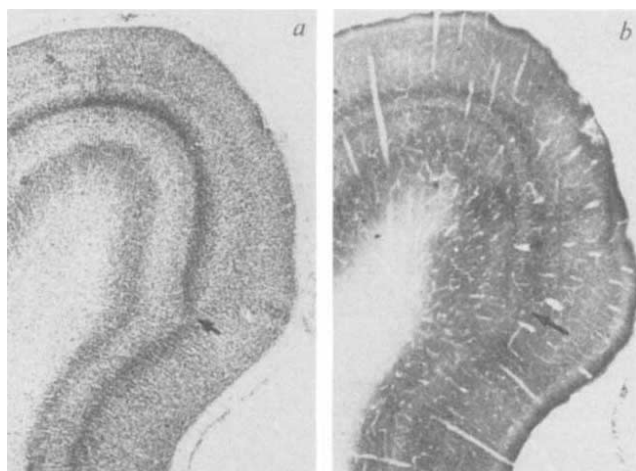


Fig. 1 Characteristics of the 17/18 border after early eye removal (E77). *a*, Section stained with cresyl violet *b*, Adjacent section reacted to show cytochrome oxidase activity. The arrows indicate the 17/18 border. Although the level of cytochrome oxidase activity is lower than in normal neonates the overall distribution in area 17 and the change in layer 4 at the 17/18 border is similar.

Methods. Timed pregnant monkeys were prepared for surgery under ketamine (i.m.) followed by Alfatesin (i.v.) anaesthesia. After intubation, anaesthesia was continued with halothane in a N₂O/O₂ mixture (70:30). Expired CO₂ and heart rate were monitored. The body temperature was maintained using a thermostatically controlled heating blanket. A midline abdominal incision was made and uterotomy performed. After exposure of the fetal head and bilateral eye removal, the fetus was replaced in the uterus and incisions closed using routine procedures. The mother was returned to her cage and medicated for the first two days with analgesic (Visceralgine, i.m.). A muscular relaxant (Duvadilan, i.m.) was given twice daily until the fetus was delivered by caesarian section 1-2 weeks before term. The infant monkey was bottle fed until the axonal tracing experiments on day E159 when removal was at E77, and E164 when eye removal was at E112 (the gestation period is 165 days). Infant monkeys were anaesthetized with ketamine (i.m.) followed by Alfatesin (i.v.). Craniotomy was performed over the occipital lobe and 15-25 µl of 30-50% horseradish peroxidase (HRP) were injected using a Hamilton syringe equipped with a glass micropipette (60-120 µm tip diameter). Closely spaced injections were made over the occipital lobe so as to massively fill area 17 and the adjacent part of area 18. After 24 hours, the infant monkey was deeply anaesthetized and perfused through the heart with a mixture of 1.25% paraformaldehyde and 1.5% glutaraldehyde. Sections (60 µm thick) were cut parasagittally and one in three sections were processed for HRP histochemistry¹⁸. Intermediate sections were processed for cytochrome oxidase¹⁹ and stained for Nissl substance (cresyl violet). The border of area 17 and 18 was clearly defined in Nissl and cytochrome oxidase-reacted sections and in the hemisphere contralateral to the injection of HRP, labelled neurons were counted and allocated to either area. Counts of labelled neurons were made in area 18 from the 17/18 border to the fundus of the lunate sulcus. Individual sections were traced out and the positions of labelled cells recorded using an x-y plotter electronically coupled to the microscope stage.

by the two methods reveals a double band of higher cytochrome oxidase activity in layer 4C that stops abruptly at the 17/18 border, as in the control animal¹⁴. Despite the weaker staining for cytochrome oxidase activity in the experimental animal its distribution within area 17 was, on the whole, similar¹⁴.

In the monkey operated on day E77, the gyral pattern of the occipital lobe was unusual and the surface area of area 17 was considerably reduced (Fig. 2, see also ref. 13), particularly laterally where, in the intact animal, cortex subserving foveal and parafoveal vision is located. Later eye removal (on day E112) did not cause a noticeable reduction in the surface area of area 17 or a disturbance in the gyral pattern¹³.

In both animals after eye removal, the disposition of callosally projecting neurons with respect to the border between areas 17

Maturation and connectivity of the visual cortex in monkey is altered by prenatal removal of retinal input

Colette Dehay*, Gwynn Horsburgh, Michel Berland†, Herbert Killackey‡ & Henry Kennedy§

INSERM-Unité 94, 16 avenue Doyen Lepine, 69500 Bron, France

* MRC Anatomical Neuropharmacology Unit, Department of Pharmacology, South Parks Road, Oxford OX1 3QT, UK

† Faculté de Médecine Lyon-Nord, Hôpital Claude Bernard, Service de Gynécologie et Obstétrique, 69600 Oullins, France

‡ Department of Psychobiology and of Anatomy and Neurobiology, University of California, Irvine, California 92717, USA

In several species, the peripheral input from the eyes partly determines the pattern of interconnections between the visual areas of the two cerebral hemispheres through the fibre tract termed the corpus callosum¹⁻⁹. In the macaque monkey, the neurons projecting through the callosum are largely restricted to area 18 throughout ontogeny, whereas area 17 is characterized by few or no callosal projections¹⁰⁻¹². Here, we show that suppressing the peripheral input by prenatal removal of the eyes leads to a marked reduction in the extent of area 17, resulting in a large shift in the position of the histologically identifiable boundary between the two areas. Even so, the boundary continues to separate an area rich with callosal connections (area 18) from one poor in such projections (area 17), indicating there is no effect on the callosal connectivity of area 17. In contrast, in area 18, eye removal results in many more neurons with callosal projections than in normal animals. The results suggest that the factors that determine the parcellation of cortical areas also specify their connectivity.

Two fetal macaque monkeys (*Macaca cynomolgus*) had both eyes removed at 77 ± 2 and 112 ± 2 days after conception (days E77 and E112), and the fetuses were then replaced in the uterus. At about the normal time of birth the callosal connectivity of areas 17 and 18 was determined using horseradish peroxidase histochemistry.

The pattern of layers in area 17 seemed remarkably normal in both experimental animals. The border between areas 17 and 18, clearly defined in Nissl-stained sections, was as distinctive as in the neonatal control (Fig. 1 and ref. 13). The 17/18 border could also be detected in sections reacted for cytochrome oxidase activity (Fig. 1). Comparison of adjacent sections stained

§ To whom correspondence should be addressed.

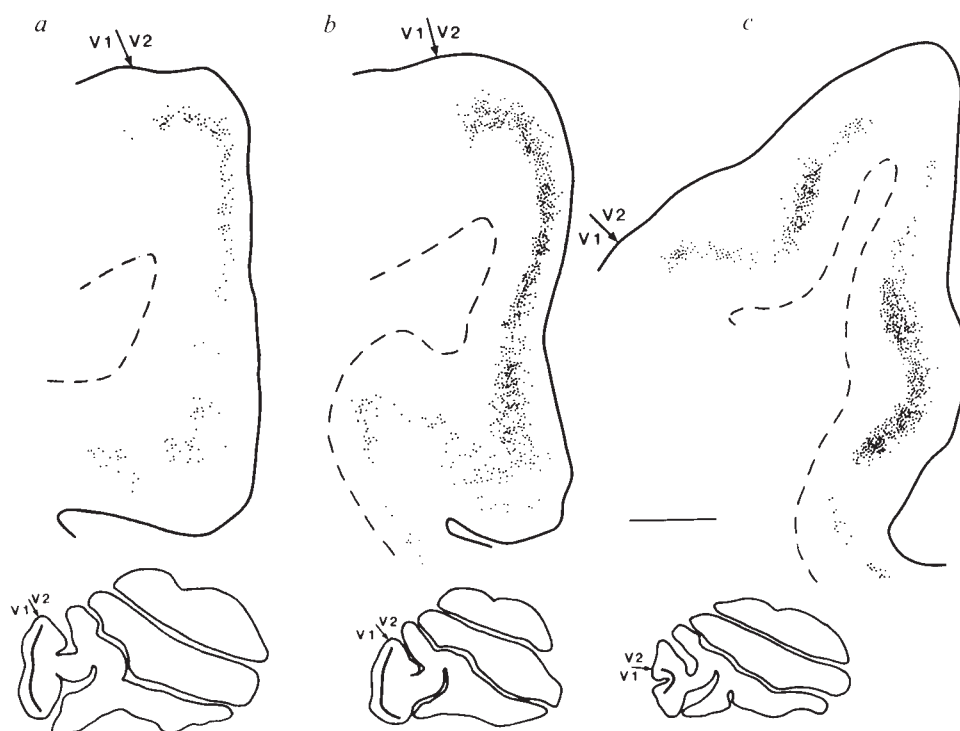


Fig. 2 Parasagittal sections of the area 17/18 border showing callosally projecting cells in the newborn monkey. *a*, Normal. *b*, Eye removal at E112. *c*, Eye removal at E77. Each dot represents a retrogradely labelled cell and arrows indicate the area 17/18 border. Inserts show low-power drawing at the level at which the high-power plots were made and where the continuous line in the operculum indicates the extent of area 17. Eye removal did not increase the numbers of callosally projecting neurons in area 17, but did increase the numbers of these cells in area 18. Whereas area 17 occupies most of the operculum of the normal animal and after eye removal at E112, after eye removal at E77, area 17 seems considerably shrunken. Scale bar, 1 mm.

and 18 was normal (Fig. 2): they were largely confined to the extrastriate cortex and stopped abruptly at the 17/18 border. In both animals, only one or two labelled neurons per section were found in area 17 where they were located, as in normal animals, within 2–3 mm of the border^{11,15,16}. In contrast, in area 18, callosally projecting neurons were more densely packed and stretched further rostrally from the 17/18 border than in normal neonates (Fig. 2). Counts of callosally projecting neurons were made in the part of area 18 that extends from the 17/18 border to the fundus of the lunette sulcus, in the part of the cortex that represents the parafoveal region of the visual field. In the normal animal there were 50–140 labelled neurons per section^{10,12}, whereas in both experimental animals the number was 5–6 times higher (280–875) than in either normal neonates or adults (Fig. 3). Although the age at which the eyes were removed did not influence the numbers of callosally projecting neurons in area 18, it did affect their spatial distribution. The animal operated on day E112 showed a continuous distribution of callosally projecting neurons but in the animal operated on day E77 the distribution of callosally projecting neurons was discontinuous and patchy (Fig. 2).

The demonstration that eye removal alters the callosal connectivity of area 18 but not of area 17 should be considered in the light of the ontogeny of callosal connections in the monkey. Several studies have shown that area 17 is largely free of callosal connections throughout development (ref. 11; L. M. Chalupa and H. Killackey manuscript, in preparation; P. Rakic, personal communication). Eye removal, which stabilizes transient callosal connections in all other species examined^{1–9}, in monkey fails to reveal a transient callosal projection in area 17, emphasizing that the lack of callosal connections in area 17, unique to the monkey, is specified early in development. The normal maturation of the cortical connectivity of area 18 differs from that of area 17, as the distribution of callosally projecting neurons in area 18 changes during development¹¹. Because eye removal both early and late in prenatal life leads to higher numbers of callosally projecting neurons in area 18, the elimination of the transient callosal projection, which normally occurs in the last month of gestation¹¹ must be regulated by epigenetic factors that depend at least partially on peripheral input.

The reduced size of area 17 after early eye removal could result from an increase in cell death. As the number of afferent

fibres from the lateral geniculate nucleus in the thalamus is reduced after early eye removal (ref. 17; H. Kennedy, unpublished observation), the survival of neurons in area 17 may depend critically on the density of the input. It is also possible that the input from the thalamus contributes directly to determining cortical areas, so that the smaller number of afferents from the lateral geniculate nucleus after eye removal leads to reduction in the extent of cortex specified as area 17 and a consequent expansion of the adjacent area 18¹³. This raises the question of whether the cortex immediately adjacent to area 17 constitutes

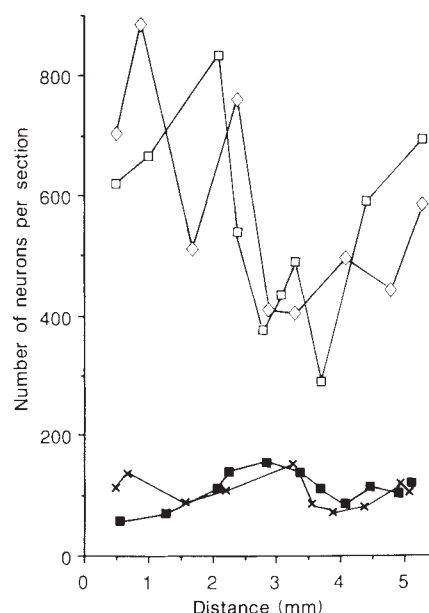


Fig. 3 Counts of HRP labelled callosally projecting neurons in area 18 of normal adult (■), neonatal control (×), after eye removal at E77 (□) and after eye removal at E112 (◇). In normal development, the numbers of callosally projecting neurons do not change after birth. Counts of callosally projecting neurons were considerably higher in the two animals whose eyes were removed. The age at which the eyes were removed had no influence on the numbers.

a hybrid area, that is, a region of cortex with characteristics of both areas 17 and 18¹³. As the absence of callosal connections from area 17 is determined early in development, it can be used to indicate whether tissue originates developmentally from area 17 or area 18. As callosally projecting neurons are found in area 18 immediately adjacent to area 17, the expanded area 18 seems to be a cortical region that was unspecified when the eyes were removed and, because of the subsequent lack of peripheral input, failed to mature into area 17.

These results indicate that retinal input has several, somewhat independent effects on cortical development. Early in development the factors that help to specify the limits of cortical visual areas also regulate the overall presence or absence of callosal connectivity. Once the boundaries of areas are established, the influence of peripheral input is limited to regulating the density of callosal projection neurons in extrastriate visual cortex.

We thank P. Rakic, S. Faugier-Grimaud, R. W. Guillery and J. Bullier for their comments on an earlier draft of this paper, and Noelle Boyer, Pascale Giroud and Sandrine Beaumont for technical assistance. This work was funded by DRET and MRES. C. D. received travel support from an ETP training grant; G. H. is supported by a Medical Foundation Research Scholarship from the University of Sydney.

Received 27 October; accepted 2 December 1988.

- Innocenti, G. M. & Frost, D. O. *Nature* **280**, 231-233 (1979).
- Innocenti, G. M. & Frost, D. O. *Expl Brain Res.* **39**, 365-375 (1980).
- Rhoades, R. W. & Dellacrose, D. D. *Brain Res.* **202**, 189-195 (1980).
- Cusick, C. G. & Lund, R. D. *J. comp. Neurol.* **212**, 385-398 (1982).
- Rothblat, L. A. & Hayes, L. L. *Brain Res.* **246**, 146-149 (1982).
- Rhoades, R. W. & Fish, S. E. *Expl Brain Res.* **51**, 451-462 (1983).
- Lund, R. D., Chang, F.-L.F. & Land, P. W. *Dev Brain Res.* **14**, 139-142 (1984).
- Grigoriadis, A. M., Murphy, E. H. & Ostrach, L. H. *Soc. Neurosci. Abstr.* **10**, 53 (1984).
- Olavarria, J. & Van Sluyters, R. C. *J. comp. Neurol.* **230**, 249-268 (1984).
- Dehay, C., Kennedy, H. & Bullier, J. *J. comp. Neurol.* **254**, 20-33 (1986).
- Dehay, C., Kennedy, H., Bullier, J. & Berland, M. *Nature* **331**, 348-350 (1988).
- Dehay, C. & Kennedy, H. *Behav. Brain Res.* **29**, 237-244 (1988).
- Rakic, P. *Science* **241**, 170-176 (1988).
- Kennedy, H., Bullier, J. & Dehay, C. *Expl Brain Res.* **61**, 204-209 (1985).
- Kennedy, H. & Dehay, C. *Behav. Brain Res.* **29**, 225-236 (1988).
- Kennedy, H., Dehay, C. & Bullier, J. *J. comp. Neurol.* **247**, 398-415 (1986).
- Rakic, P. & Williams, R. W. *Soc. Neurosci. Abstr.* **12**, 1499 (1986).
- Mesulam, M. M. *et al. J. Histochem. Cytochem.* **28**, 1255-1259 (1980).
- Wong-Riley, M. *Brain Res.* **171**, 11-28 (1979).

Effect of recombinant soluble CD4 in rhesus monkeys infected with simian immunodeficiency virus of macaques

Mamoru Watanabe, Keith A. Reimann, Peter A. DeLong, Theresa Liu*, Richard A. Fisher*, & Norman L. Letvin†

Harvard Medical School, New England Regional Primate Research Center, Southborough, Massachusetts 01772, USA

* Biogen Research Corporation, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

The CD4 molecule is a high-affinity cell-surface receptor for the human immunodeficiency virus (HIV-1)¹ and a soluble truncated form of CD4 produced by recombinant DNA technology is a potent inhibitor of HIV-1 replication and HIV-1-induced cell fusion *in vitro*²⁻⁶. Rhesus monkeys infected with the simian immunodeficiency virus of macaques (SIV_{MAC}), a virus closely related to HIV-1, develop an AIDS-like syndrome, and so provide an important model for the evaluation of potential AIDS therapies⁷. We have assessed the therapeutic effect of recombinant soluble CD4 in SIV_{MAC}-infected rhesus monkeys. Virus was readily isolated from peripheral blood lymphocytes and bone marrow cells of these animals before starting treatment with

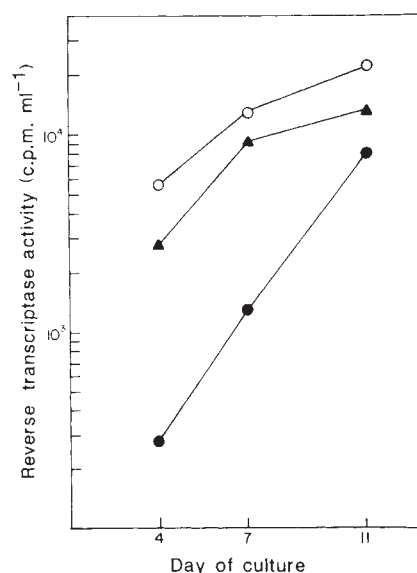


Fig. 1 Inhibition by rsCD4 of SIV_{MAC} infection of normal rhesus monkey PBLs *in vitro*. SIV_{MAC}-containing culture supernatant was incubated for 1 h at 37 °C with culture medium alone or with two concentrations of rsCD4 and then added to 2.5×10^6 cell aliquots of concanavalin A activated normal rhesus monkey PBLs in 1 ml, resulting in final concentrations of no rsCD4 (○), $12.5 \mu\text{g ml}^{-1}$ rsCD4 (▲), and $125 \mu\text{g ml}^{-1}$ rsCD4 (●). After a 2-h incubation at 37 °C, these cells were brought up to 5 ml in IL-2-supplemented medium. Following an overnight incubation, the cell populations were washed, resuspended in IL-2-supplemented medium at 5×10^5 cells per ml and maintained at 37 °C in a humidified 5% CO₂ atmosphere. Culture medium was collected every 3-4 days and assayed for reverse transcriptase activity¹⁴.

soluble CD4, but became difficult to isolate soon after treatment had begun. Moreover the diminished growth of both granulocyte-macrophage and erythrocyte progenitor colonies from the bone marrow of these monkeys rose to normal levels during treatment. These findings indicate that soluble CD4 could prove valuable in the treatment of AIDS.

Our growing understanding of HIV-1 and its interaction with the cells it infects is leading to rational strategies for AIDS therapy. The CD4 molecule acts as a high-affinity receptor for the outer viral envelope glycoprotein (gp120) of HIV-1¹. Monoclonal anti-CD4 antibodies block the ability of HIV-1 to infect target cells *in vitro*^{8,9}, and the transfer and expression of the CD4 gene to previously CD4⁻ human cells confers susceptibility to HIV-1 infection on those cells¹⁰. It has recently been shown that a recombinant soluble form of the CD4 molecule (rsCD4) is a potent inhibitor of both HIV-1 replication and virus-induced cell fusion *in vitro*²⁻⁶. These findings led to the proposal that rsCD4 may be useful in AIDS therapy.

Lentiviruses related in their nucleotide sequences to the human AIDS viruses HIV-1 and HIV-2 have recently been isolated from African and Asian non-human primates¹¹. Some of these isolates, when inoculated into rhesus monkeys, induce a disease with striking similarities to human AIDS^{7,12,13}. The SIV_{MAC}-induced disease in the rhesus monkey has already proved to be an important model for studying the pathogenesis of AIDS and for assessing strategies for vaccination and therapeutic interventions in this disease^{14,15}. The human and macaque CD4 molecules are strikingly similar both in their monoclonal anti-CD4 epitope mapping and in their amino-acid sequences that define the minimal region for binding the outer envelope glycoprotein of the AIDS virus (R.A.F. *et al.*, unpublished data). In agreement with these observations, we find that a recombinant soluble human CD4, a 375-amino acid protein comprising the entire extracellular portion of the molecule², blocks rhesus monkey peripheral blood lymphocyte (PBL) infection by SIV_{MAC} *in vitro* (Fig. 1). Although human rsCD4 may

† To whom correspondence should be addressed.

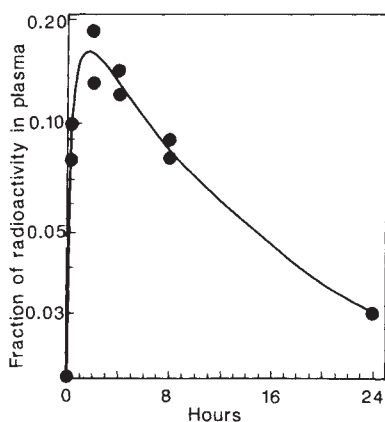


Fig. 2 Plasma clearance of total radioactivity from 2 rhesus monkeys following intramuscular administration of ^{125}I -labelled rsCD4. Radiolabelled rsCD4 (1.1×10^7 c.p.m.) was administered with 2 mg unlabelled rsCD4, i.m., into each of 2 normal 6.9 kg rhesus monkeys. Samples of plasma (1 ml) were regularly analysed for radioactivity. The fraction of the total initial radioactivity in the plasma was calculated assuming a plasma volume of 40 ml kg^{-1} .

be less efficient in blocking SIV_{MAC} infection of rhesus monkey PBLs than HIV-1 infection of human PBLs, the SIV /rhesus monkey model provides a reasonable system in which to assess the therapeutic potential of rsCD4 in AIDS.

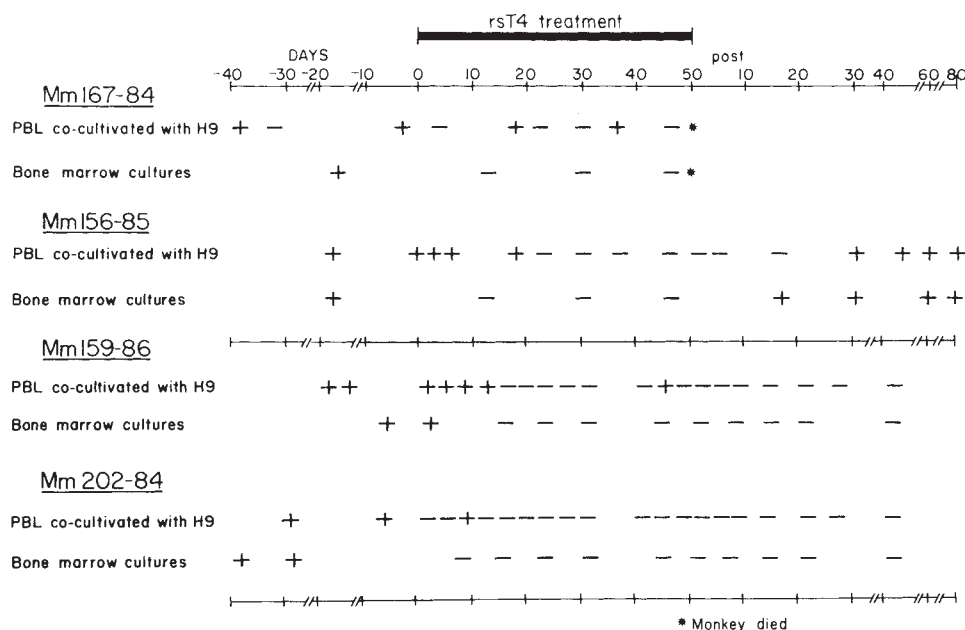
Pharmacokinetic studies were performed to establish an appropriate route and frequency of administration of rsCD4 to the rhesus monkey. Animals weighed 5–7 kg. Two normal rhesus monkeys were given a single 2 mg intramuscular (i.m.) inoculation of rsCD4 with a trace quantity of ^{125}I -labelled rsCD4 admixed. As shown in Fig. 2, >10% of the calculated maximum possible fraction of radioactivity was found in plasma 15 min after a single dosage and persisted for 8 hours. Peak plasma levels of $>1,400 \text{ ng ml}^{-1}$ of rsCD4 were measurable in an enzyme-linked immunosorbent assay two hours after its administration. This method of delivery yielded an estimated rsCD4 serum half life of about 6 hours. Two uninfected rhesus monkeys were then given a 50-day course of rsCD4 treatment, each receiving 2 mg, i.m., daily, and did not experience any significant adverse reaction to the rsCD4. The distribution of lymphocyte subsets in their peripheral blood (CD2^+ , CD4^+ , CD8^+ , IL-2-receptor^+ , B1^+) and the function of those cells

(proliferation following stimulation with xenogeneic cells and pokeweed mitogen) did not change during the course of treatment. Serum chemistry studies to assess liver and renal function on days 8, 15 and 44 of treatment and on days 8 and 18 after treatment were unchanged when compared with baseline values. The animals did experience a transient drop in the number of bone marrow granulocyte-macrophage progenitor (CFU-GM) colonies with an associated transient neutropenia, but this transient bone marrow suppression never became clinically significant.

Four SIV_{MAC} -infected rhesus monkeys were then treated with rsCD4 in the same way. Three of these virus-infected monkeys had nearly normal immune parameters at the start of treatment; the fourth had a circulating absolute CD4^+ lymphocyte count of less than 100 per mm^3 (normal rhesus monkeys have $>1,000 \text{ CD4}^+$ PBLs per mm^3). None of the SIV_{MAC} -infected monkeys showed changes during treatment in the distribution of lymphocyte subsets in their peripheral blood, in PBL function or in serum chemistry assessment of liver and renal function. The SIV_{MAC} -infected animal (Mm167-84) that showed evidence of severe immunological impairment at the start of the study died on the third day after the end of treatment. There was evidence of a severe Coombs positive haemolytic anaemia. In the last 10 days before death, the animal's haemoglobin dropped from 7.7 to 3.7 g dl^{-1} . Lesions seen at necropsy included a large amount of clotted blood in the small intestine, severe myocardial fibrosis, and histiocytic infiltration of the splenic lymphoid follicles. These last two findings are not uncommon in rhesus monkeys chronically infected with SIV_{MAC} . The presence of blood in the proximal small bowel indicated that the animal could have died from an acute gastrointestinal haemorrhage, which was possibly secondary to stress induced by the acute haemolytic process. As significant haemolytic anaemias have been reported in AIDS¹⁶, it is impossible to determine whether this death was a natural consequence of the SIV_{MAC} infection or was drug-induced.

The results of the virus-isolation studies in these four monkeys were striking (Fig. 3). Co-cultivation of PBLs from animals Mm156-85, Mm159-86 and Mm202-84 with H9 yielded SIV_{MAC} before treatment and on every isolation attempt up to about the end of week 2 of treatment. These virus isolation attempts thereafter were negative until one month after completion of treatment at the earliest. SIV_{MAC} replication in bone marrow cultures, a reflection of virus replication in macrophages rather than in T lymphocytes (data not shown) showed parallel

Fig. 3 SIV_{MAC} could not be isolated from PBLs and bone marrow of the virus-infected monkeys during treatment with rsCD4. PBLs of SIV_{MAC} -infected animals were isolated by Ficoll-diatrizoate density gradient ($\rho = 1.077$) centrifugation, activated with concanavalin A and co-cultured with the human T-cell line H9. Bone-marrow samples were obtained by posterior iliac crest aspiration from ketamine anesthetized monkeys and heparinized. Cells were isolated by density gradient centrifugation, washed and placed in culture in Iscove's modified Dulbecco's MEM supplemented with 12.5% FBS and 12.5% horse serum in 2 or 4 chamber Lab-Tek tissue culture slides (Miles Scientific) at a cell concentration of 1×10^6 per ml. After 7 days of culture, non-adherent cells were removed. In both culture systems, medium was changed regularly and supernatants were assayed for reverse transcriptase activity as previously described¹⁴. Supernatants with reverse transcriptase activity at least 5 times above background were scored as positive for infection.



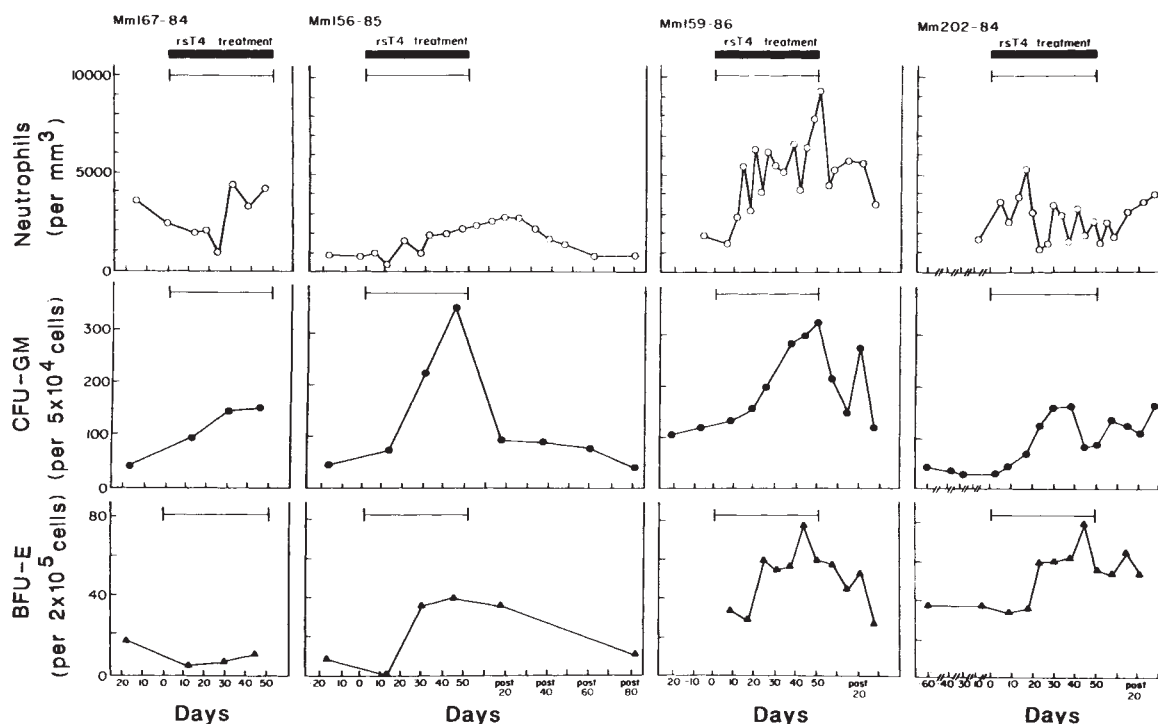


Fig. 4 SIV_{MAC}-infected rhesus monkeys demonstrate an increase in bone marrow granulocyte-macrophage and erythrocyte progenitor cell growth following rsCD4 treatment. A two-layer culture was established for the quantitation of CFU-GM colonies from bone marrow cells isolated as described in Fig. 3. The underlayer of 0.5% Noble agar contained 60 ng ml⁻¹ recombinant human granulocyte-macrophage colony stimulating factor (kindly provided by Genetics Institute, MA). The overlay contained 10⁵ bone marrow cells in Iscove's modified DMEM supplemented with 12.5% FBS and 12.5% horse serum in 0.3% Noble agar. It was layered on to the underlayer in 35 × 10 mm tissue-culture dishes and the cultures were maintained at 37 °C in a 5% CO₂-humidified atmosphere. BFU-E colonies were assessed in cultures of bone marrow cells maintained in 0.9% (w/v) methylcellulose in Iscove's modified DMEM supplemented with 30% FBS, 0.9% deionized bovine serum albumin and 5 × 10⁻⁵ M 2-mercaptoethanol, containing 60 ng ml⁻¹ granulocyte-macrophage colony stimulating factor, 1 unit ml⁻¹ sheep erythropoietin (step III, Connaught Lab., Ontario) and 5% phytohaemagglutinin-stimulated rhesus monkey lymphocyte conditioned medium. CFU-GM and BFU-E colonies of >50 cells were then counted under an inverted microscope 12–14 days after establishing the cultures. Total neutrophil counts in the peripheral blood were determined from complete blood counts and differentials from EDTA-anticoagulated blood samples.

changes. Virus was readily isolated from bone marrow before treating these 3 monkeys, but attempts to isolate virus during treatment were negative. Virus isolations from bone marrow reverted to positive on day 18 after completion of treatment in Mm156-85, but remained negative until 43 days after treatment in Mm159-86 and Mm202-84. Virus was isolated only episodically from PBLs of the SIV_{MAC}-infected rsCD4-treated animal Mm167-84 following co-cultivation with H9 cells both before and during treatment (Mm167-84 was the animal that was immunologically compromised at the outset of the study and died 3 days after completion of treatment). But although isolation of virus from the bone marrow of Mm167-84 was positive before treatment, 3 attempts to isolate virus from the bone marrow during treatment were unsuccessful.

These negative virus isolations alone were not unequivocal evidence for a therapeutic effect of rsCD4; the rsCD4 could simply have interfered with the *in vitro* isolation of SIV_{MAC} without affecting virus replication or the sequelae of these infections in the monkeys. But the fact that virus isolations remained negative in three animals for longer than a month after treatment argues against this possibility. Moreover, functional abnormalities reflecting SIV_{MAC}-induced disease also improved in these animals. The *in vitro* growth of both CFU-GM and erythrocyte progenitor (BFU-E) colonies from the bone marrow of these monkeys was depressed at the start of rsCD4 treatment. This is a common finding in SIV_{MAC}-infected monkeys (data not shown) and in HIV-1-infected humans¹⁷. (CFU-GM colony counts in a series of studies of 10 normal animals were 222 ± 16 per 5 × 10⁴ cultured bone marrow cells; BFU-E counts were 76 ± 4 per 2 × 10⁵ cultured bone marrow cells.) The number of colony-forming cells from the bone marrow of these animals

increased during rsCD4 treatment (Fig. 4). A bone marrow cell CFU-GM colony count of 48 per 5 × 10⁴ cells before treatment of Mm167-84 rose during treatment to 146 per 5 × 10⁴ cells with no significant change in the number of BFU-E colonies. The CFU-GM colony count derived from bone marrow of Mm156-85 rose from 44 per 5 × 10⁴ cells before the start of therapy to 350 per 5 × 10⁴ cells, with a coincident rise in BFU-E colonies from 8 to 40 per 2 × 10⁵ cells. Bone marrow from Mm159-86 showed a rise in CFU-GM colonies from 106 to 329 per 5 × 10⁴ cells and in BFU-E colonies from 34 to 78 per 2 × 10⁵ cells. Bone marrow from Mm202-84 gave an increase in CFU-GM colonies from 46 to 163 per 5 × 10⁴ cells and in BFU-E colonies from 38 to 80 per 2 × 10⁵ cultured cells. A decline in bone marrow CFU-GM and BFU-E colony counts to baseline values occurred in the three surviving animals after stopping treatment. A transient neutropenia occurred in two of the four monkeys during the first two weeks of treatment. But there was an eventual rise in total peripheral blood neutrophil count in all four animals that accompanied the increased colony-forming cell count from the bone marrow.

The rsCD4 treatment of the two normal rhesus monkeys unexpectedly had an anti-viral effect. More than 90% of bone marrow cultures established from uninfected monkeys can be reproducibly infected with exogenously introduced SIV_{MAC} (data not shown). But although bone marrow from both the rsCD4-treated normal animals was susceptible *in vitro* to infection with SIV_{MAC} introduced before and 4 weeks after treatment, 3 attempts to infect bone marrow cell cultures from the animals during the 50-day treatment period and a single attempt 18 days afterwards gave negative cultures.

Our inability to isolate virus from PBLs of the SIV_{MAC}-

infected animals by H9 co-cultivation during rsCD4 treatment did not mean that these animals were cleared of virus. If CD8⁺ lymphocyte-depleted rhesus monkey PBLs are co-cultivated with human PBLs in IL-2-supplemented medium, low levels of virus infection can be detected¹⁶. Using this method we were able to isolate virus from these animals during treatment. Nevertheless virus was more difficult to isolate during treatment than before and after treatment, and this probably reflects a quantitative decrease in replication of SIV_{MAC} in these animals. The coincident improvement in bone marrow function during this period of treatment is a functional benefit probably related to the decreased virus load. This increase in bone marrow colony-forming cells was not seen in the treated normal animals. The unexpected protection of bone marrow cultures from treated normal animals against SIV_{MAC} infection suggests another therapeutic benefit of rsCD4 treatment.

It has been suggested that rsCD4 could act by absorbing soluble gp120 and thus interfering with cytolytic T lymphocyte killing of HIV-1 infected targets, or by inhibiting cell fusion, or by direct blocking of virus-target cell binding¹⁹. It could also act indirectly to provide a therapeutic benefit to HIV-1-infected individuals. Although we do not know the precise mechanism by which it works, the beneficial effect of the rsCD4 treatment in SIV_{MAC}-infected rhesus monkeys suggests that it could be of value in treatment of humans infected with HIV-1.

We thank N. King for pathology; C. Lord, K. Solomon, S. Turner, M. Stone and B. Burrus for technical assistance; M. Evans for help with animal treatments; J. Maragarore, W. Meier and the Biogen Process Development Group for preparation of the rsCD4; V. Johnson for assistance with the Coombs assay; R. Schooley and V. Sato for discussion and D. Brosseau for preparing the manuscript. This work was supported by Biogen, Inc. and the NIH.

Received 14 October; accepted 8 December 1988.

1. McDougal, J. S. *et al.* *Science* **231**, 382-385 (1986).
2. Fisher, R. A. *et al.* *Nature* **331**, 76-78 (1988).
3. Hussey, R. E. *et al.* *Nature* **331**, 78-81 (1988).
4. Deen, K. C. *et al.* *Nature* **331**, 82-84 (1988).
5. Trautner, A., Luke, W. & Karjalainen, K. *Nature* **331**, 84-86 (1988).
6. Smith, D. H. *et al.* *Science* **238**, 1704-1707 (1987).
7. Letvin, N. L. *et al.* *Science* **230**, 71-73 (1985).
8. Dalgleish, A. G. *et al.* *Nature* **312**, 763-766 (1984).
9. Klatzmann, D. *et al.* *Nature* **312**, 767-768 (1984).
10. Maddon, P. J. *et al.* *Cell* **47**, 333-348 (1986).
11. Letvin, N. L. in *AIDS* (ed. Kulstad, R.) (AAAS, Washington D.C., 1989).
12. Fultz, P. N. *et al.* in *Vaccines 88* (ed. Chanock, R. M.) (Cold Spring Harbor Laboratory, New York 1988).
13. Murphy-Corb, M. *et al.* *Nature* **321**, 435-437 (1986).
14. Kannagi, M. *et al.* *J. clin. Invest.* **78**, 1229-1236 (1986).
15. Letvin, N. L. *et al.* in *Vaccines 87* (ed. Chanock, R. M.) (Cold Spring Harbor Laboratory, New York 1987).
16. Schreiber, Z. A., Loh, S. H., Charles, M. & Abebe, L. S. *Blood* **52**, 117a (1983).
17. Stella, C. C., Ganser, A. & Hoelzer, D. *J. clin. Invest.* **80**, 286-293 (1987).
18. Kannagi, M., Chalifoux, L. V., Lord, C. I. & Letvin, N. L. *J. Immun.* **140**, 2237-2242 (1988).
19. Fauci, A. S. *Science* **239**, 617-622 (1988).

Induction of endogenous IFN- α and IFN- β genes by a regulatory transcription factor, IRF-1

Takashi Fujita, Yoko Kimura, Masaaki Miyamoto, Edward L. Barsoumian & Tadatsugu Taniguchi

Institute for Molecular and Cellular Biology Osaka University, Suita-shi, Osaka 565, Japan

Interferons (IFNs) have an important role in cell growth and differentiation. The most well-known function of IFNs is their antiviral activity; viral infections result in induction of the transcription of the IFN- α and IFN- β genes¹⁻³. Recently we isolated the gene encoding a transcription factor, IRF-1, that may play a part in the induction of IFN genes^{4,5}. Interestingly, the IRF-1 gene itself is virus-inducible, suggesting the importance of *de novo*

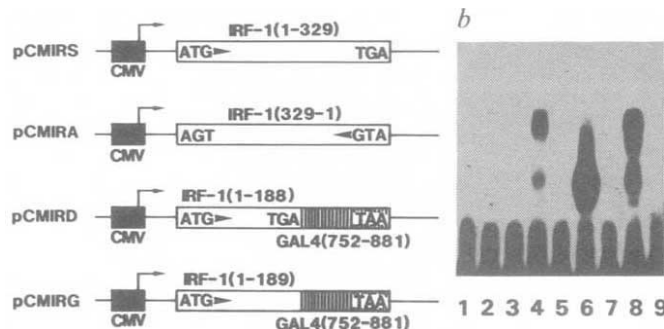


Fig. 1 Expression of murine IRF-1 and its derivatives in COS cells. **a**, Structure of expression vectors. Numbers in parentheses indicate the amino-acid residues of each factor. In all constructs the genes were inserted into the expression vector CDM8, which contains the cytomegalovirus (CMV) promoter/enhancer and the SV40 *ori* sequences¹³. They were (1) pCMIRS, which contains the entire mouse IRF-1 coding sequence downstream of the CMV promoter/enhancer sequence in the sense orientation; (2) pCMIRA, which is the same as pCMIRS except that the IRF-1 gene is introduced in the antisense orientation; (3) pCMIRD, in which a translational terminator sequence has been introduced within the cDNA sequence of pCMIRS, resulting in a gene product that lacks the carboxyl-terminal portion of the IRF-1 (that is, downstream from amino acid 189); (4) pCMIRG, which encodes a chimaeric factor consisting of the amino-terminal portion of IRF-1 and the yeast GAL4 transcriptional activation domain¹⁴. **b**, Detection of DNA-binding activity in COS cell transfectants. DNA transfection was performed by a method modified from Seed and Aruffo¹⁵. COS cells in a 10-cm dish (5×10^6 cells per dish) were transfected in 5 ml Iscove's modified Eagle's medium containing 10% NuSerum, 500 $\mu\text{g ml}^{-1}$ DEAE dextran, 100 μM chloroquin diphosphate and 5 $\mu\text{g ml}^{-1}$ plasmid DNA. Transfection efficiency of this protocol is $\sim 30\%$, as estimated by the Tac antigen (IL-2 receptor)-positive COS cell population after transfection of pSVIL2R-3 (ref. 16). Whole cell extracts of transfected COS cells with pCMIRA (lanes 2 and 3), pCMIRS (lanes 4 and 5), pCMIRD (lanes 6 and 7) and pCMIRG (lanes 8 and 9) were prepared 48 h after transfection⁴. Each extract was subjected to gel shift analysis using ³²P-labelled C13 oligomer (sequence AAGTGA repeated 3 times in tandem) as probe⁶. The binding mixture (10 μl) contains 20 mM HEPES, pH 7.6, 0.5 mM dithiothreitol, 0.1 mM EGTA, 1 mM MgCl_2 , 4% Ficoll, herring sperm DNA (1 μg), poly(dI-dC)·poly(dI-dC) (2 μg), ³²P-probe (5 fmole) and 2 μg protein of whole-cell extract. In lanes 3, 5, 7 and 9, a molar excess of unlabelled C1 oligomer (AAGTGA repeated 4 times in tandem)⁶ (5 pmole) is also included. Lane 1: no extract. A low-level of the monkey IRF-1-like activity can be seen in lane 2. The shifted bands migrating faster in lanes 4 and 8 might be breakdown products of IRF-1 which still retain the DNA-binding property of native IRF-1. The intensity of the shifted bands was about 8 (pCMIRS), 110 (pCMIRD) and 32 (pCMIRG) relative to the control (pCMIRA). Following viral induction, a twofold increase is observed in untransfected controls (data not shown).

production of IRF-1 in IFN gene induction⁵. Here we show that high-level expression of the cloned mouse IRF-1 gene in monkey COS cells results in the induction of endogenous IFN- α and IFN- β genes without viral stimulation. Furthermore, we demonstrate the induction of these genes by an IRF-1/yeast GAL4 chimaeric transcription factor. This may be the first demonstration of the specific induction of silent chromosomal genes by transfection of a single transcription factor gene in mammalian cells.

In order to study the function of IRF-1 in IFN gene expression, we constructed a series of expression vectors for the mouse IRF-1 gene⁵ (Fig. 1a). The expression vectors were each transfected into monkey COS cells and the gene products were analysed by gel retardation using cell extracts and a ³²P-labelled C13 oligomer probe⁶. As shown in Fig. 1b, a significant increase in DNA binding activity was observed in the transfectants pCMIRS, pCMIRD and pCMIRG, demonstrating that the expected products are in fact produced in large amounts (see legend).

We next tested whether the increased IRF-1 expression results in the induction of otherwise silent IFN genes. As shown in

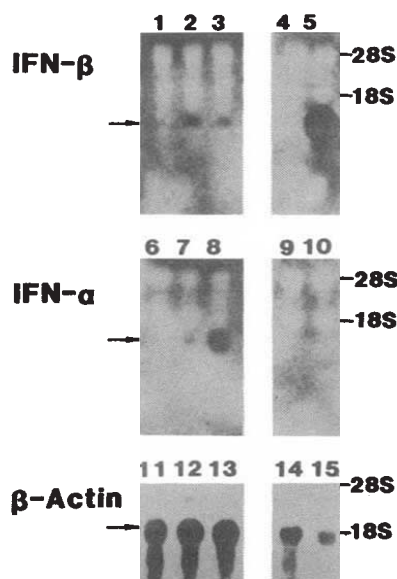


Fig. 2 Detection of IFN- β and IFN- α I specific messenger in COS cell transfectants. Twenty μ g poly(A)⁺ RNA from COS cells transfected with pCMIRA (lanes 1, 6 and 11), pCMIRS (lanes 2, 7 and 12) and pCMIRG (lanes 3, 8 and 13) and 10 μ g total RNA from mock-infected COS cells (lanes 4, 9 and 14) and NDV-infected (10 h) COS cells (lanes 5, 10 and 15) were subjected to northern blotting analysis. It has been shown that cross-hybridization analysis of monkey genomic DNA with human IFN- α and IFN- β probes results in very similar patterns under stringent hybridization conditions¹⁷. Hybridization was with a ³²P-labelled human IFN- β probe (1.6-kilobase (kb) *Bam*HI-*Eco*RI fragment of pSE-66 (ref. 18), 8×10^8 c.p.m. μ g⁻¹, lanes 1-5), human IFN- α I probe (0.7 kb *Hgi*AI fragment of Hif-2h cDNA¹⁹, 9.2×10^8 c.p.m. μ g⁻¹, lanes 6-10) and human β -actin probe (2.0 kb *Bam*HI-*Pvu*II fragment of λ Ha-204 (ref. 5), 8×10^8 c.p.m. μ g⁻¹, lanes 11-20). The filters were exposed at -80 °C for 2 days (IFN- α and IFN- β) or 4 h (actin). Densitometric measurement revealed the relative intensities of the bands in lanes 2, 3, 5, 7 and 8 as 1, 1, 259, 1 and 8.7, respectively. The apparent increase in IFN- α mRNA in lane 8 possibly represents induction of multiple IFN- α genes, the product of which might not have the same specific activity². 28S and 18S ribosomal RNA size markers are shown to the right.

Table 1, antiviral activity was detectable in culture supernatants of the COS cells transfected with pCMIRS and pCMIRG but not with pCMIRA and pCMIRD. The detected IFN activity was attributable to IFN- α I, IFN- α II (also referred to as IFN- ω ; ref. 7) and IFN- β . The IFN- β production was also confirmed by enzyme-linked immunosorbent assay. A combination of antibodies specific for human IFN- α I, α II and β completely neutralized the antiviral activity, implying that type II IFN (IFN- γ) was not generated using the transfectants. In contrast, COS cells infected with Newcastle disease virus (NDV) produced primarily IFN- β (Table 1). Induction of IFN mRNA was analysed by northern blotting. Both IFN- α and IFN- β messenger RNAs were detected specifically in the pCMIRS- and pCMIRG-, but not in pCMIRA- and pCMIRD-transfected COS cells. Although IRF-1 displays a lower affinity for IFN- α I compared with IFN- β (ref. 5), comparable levels of mRNA were induced. The overall IFN- α mRNA may be a reflection of the multiple subtypes induced². The levels of IFN- β mRNA were much lower when compared with NDV-induced COS cells, but the mRNA size was the same in all samples, suggesting that mRNA initiation had occurred at the genuine CAP site of each gene. Similar results were obtained by expressing a complementary DNA encoding the human IRF-1 (data not shown). From the data, it seems that the amino-terminal region of IRF-1 is responsible for recognition of the regulatory DNA sequences, whereas the carboxyl-terminal region participates in activating

Table 1 IFN production in COS cells

COS cell supernatant	Interferon activity (U/5 $\times 10^6$ cells)	
	Experiment 1	Experiment 2
Virus-induced cells		
Mock-induced (control)	<1	ND
NDV-induced*	5,120	ND
NDV-induced + anti-Namalwa-IFN†	5,120	ND
NDV-induced + anti-IFN- β ‡	<1	ND
Antiviral activity in DNA-transfected cells		
pCMIRA	<1	<1
pCMIRS	160	160
pCMIRD	<1	<1
pCMIRG	320	230
pCMIRG + anti-Namalwa IFN	ND	80
pCMIRG + anti-IFN- α II§	ND	60
pCMIRG + anti-IFN- β	ND	110
pCMIRG + anti-Namalwa, α II and β	ND	<1
IFN-β specific ELISA¶		
pCMIRD	ND	<1
pCMIRG	ND	14

Values are standardized against reference human IFN- β . ND, not determined.

* Virus inactivated at pH 2.

† Anti-Namalwa IFN- α serum (polyclonal) (from Dr S. Yonehara, Tokyo Metropolitan Institute of Medical Science), final concentration 1%. The purified Namalwa-derived IFN used for the antibody preparation consists mainly of IFN- α I, but the presence of IFN- α II and IFN- β are not strictly ruled out (ref. 12; S. Yonehara, personal communication). Therefore, the IFN- α I level measured by the neutralization experiment could be an overestimation.

‡ Anti-human IFN- β serum (polyclonal) (from Dr T. Sudo, Biomedical Institute, Tokyo), final concentration 2.5%.

§ Anti-recombinant human IFN- α II serum⁷ (polyclonal) (from Dr G. Adolf, Ernst Boehringer Institut für Arzneimittelforschung, Vienna), final concentration 2.5%.

¶ Enzyme-linked immunosorbent assay (Toray, Tokyo). The value is likely to be an underestimation as the antibodies (the enzyme conjugate consists of a monoclonal antibody) are against human and not monkey IFN- β .

transcription. IFN- α mRNA seems to be preferentially induced by the chimaeric factor (Fig. 2).

Two interesting issues arise from these observations: (1) Why is the IFN- β expression level much lower in the transfected cells compared with the NDV-induced cells? (2) Why are IFN- α genes induced by transfecting the IRF-1 cDNA but not by NDV in COS cells? In addressing the first question, it is possible that the IFN- β gene is subject to control mechanism(s) that can be relieved upon viral induction^{4,8-10}. In the IRF-1 gene-transfected cells, IRF-1 produced in large amounts might overcome such a control mechanism, resulting in activation of transcription, albeit at a low level. Alternatively, an additional signal(s) that can be delivered following viral infection may be needed^{4,6}. In fact, NDV induction of pCMIRS-transfected COS cells resulted in about a 3- and 4-fold increase in induced IFN- β and IFN- α mRNA respectively (data not shown). These possibilities are not mutually exclusive. As for the second question, the molecular events causing this interesting difference are unclear at present. It is most likely that IRF-1 promotes transcription of IFN- α genes by interacting with their regulatory elements, because many, if not all, of the genes contain the elements recognized by IRF-1^{4,5,10}. It is possible that the factor binds to those elements by displacing proteins (either specific factors or non-specific proteins such as histones) bound to chromosomal IFN- α genes. In cells that can naturally produce IFN- α , for example monocytes or B cells^{1,2,11}, the chromosomal structures might be

different from IFN- β . The results presented in this study provide strong evidence that the IRF-1 regulates the expression of type I IFN genes.

We thank Dr Jan Vilček for critically reviewing the manuscript, and Drs J. Ma and M. Ptashne for RB968 and for valuable comments. The work is supported in part by a grant-in-aid for Special Project Research, Cancer Bioscience from the Ministry of Education, Science and Culture of Japan, and by the Nissan Science Foundation.

Received 10 October; accepted 30 November 1988.

1. Lengyel, P. A. *Rev. Biochem.* **51**, 251-282 (1982).
2. Weissmann, C. & Weber, H. *Nucleic Acids Res.* **33**, 251-300 (1986).
3. Taniguchi, T. A. *Rev. Immun.* **6**, 439-464 (1988).
4. Fujita, T. *et al. EMBO J.* **7**, 3397-3405 (1988).
5. Miyamoto, M. *et al. Cell* **54**, 903-913 (1988).
6. Fujita, T., Shibuya, H., Hotta, H., Yamanishi, K. & Taniguchi, T. *Cell* **49**, 357-367 (1987).
7. Adolf, G. *J. gen. Virol.* **68**, 1669-1676 (1987).
8. Zinn, K., DiMaio, D. & Maniatis, T. *Cell* **34**, 865-879 (1983).
9. Goodbourn, S., Burstein, H. & Maniatis, T. *Cell* **45**, 601-610 (1986).
10. Kuhl, D. *et al. Cell* **50**, 1057-1069 (1987).
11. Hiscott, J., Cantell, K. & Weissmann, C. *Nucleic Acids Res.* **12**, 3727-3746 (1984).
12. Shuttleworth, J., Moser, J. & Burck, D. C. *Eur. J. Biochem.* **133**, 399-404 (1983).
13. Seed, B. *Nature* **329**, 840-842 (1988).
14. Ma, J. & Ptashne, M. *Cell* **48**, 847-853 (1987).
15. Seed, B. & Aruffo, A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3365-3369 (1987).
16. Hatakeyama, M. *et al. Nature* **318**, 467-470 (1985).
17. Wilson, V. *et al. J. molec. Biol.* **166**, 457-475 (1983).
18. Fujita, T., Ohno, S., Yasumitsu, H. & Taniguchi, T. *Cell* **41**, 489-496 (1985).
19. Mantei, N. *et al. Gene* **10**, 1-10 (1980).

Phosphorylcholine acts as a Ca^{2+} -dependent receptor molecule for lymphocyte perforin

Jürg Tschopp, Sylvie Schäfer, Danièle Masson, Manuel C. Peitsch & Christoph Heusser*

Institute of Biochemistry, University of Lausanne, CH-1066 Epalinges, Switzerland

* Ciba-Geigy AG, CH-4002 Basle, Switzerland

Large granular lymphocytes and cytolytic T-lymphocytes (CTL) contain numerous cytoplasmic granules thought to be responsible, at least in part, for the cytolytic activity of these effector cells. Isolated granules are lytic for a variety of target cells and the granule proteins are specifically released upon target-cell interaction¹⁻⁴. Major proteins in mouse CTL granules are a family of seven serine proteases designated granzymes A to G⁵, and a pore-forming protein called perforin (cytolysin)^{1-4,6}. Purified perforin is cytolytic in the presence of Ca^{2+} and shows ultrastructural, immunological and amino-acid sequence similarities to complement component C9⁶⁻⁸. Despite these similarities, perforin and C9 are clearly distinct in their mode of target-cell recognition. Whereas C9 insertion is absolutely dependent on a receptor moiety assembled from the complement proteins C5b, C6, C7, and C8⁹ on the target-cell membrane, no requirement for a receptor molecule has been reported for perforin. Here, we demonstrate that phosphorylcholine acts as a specific, Ca^{2+} -dependent receptor molecule for perforin.

Although complement component C9, in the absence of the C5b-8 receptor complex, inserts into and lyses small unilamellar vesicles⁹, large multilamellar vesicles or membranes of erythrocytes are refractory to this C9 attack¹⁰. C9 is only able to intercalate non-specifically into the lipid bilayer without a receptor molecule if the density of lipid headgroups is low, as in the case of highly curved small vesicles. Similarly, perforin lyses small vesicles or other types of unilamellar vesicles without any apparent lipid specificity or need for a proteinaceous receptor molecule^{2,11,12}. To exclude the possibility of non-specific insertion into this class of vesicles however, large multilamellar vesicles were tested for their inhibitory activity of perforin-mediated

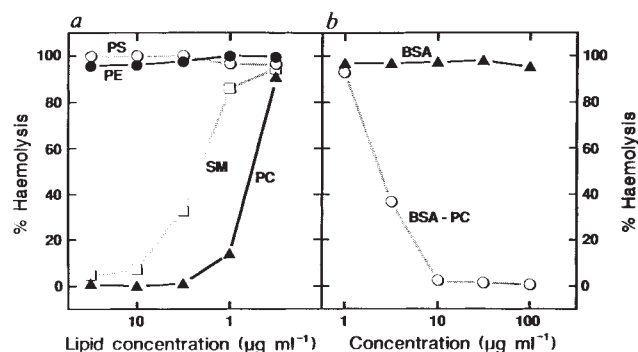


Fig. 1 Inhibition of perforin-mediated haemolysis by lipid vesicles. **a**, Inhibition of haemolysis by vesicles composed of phosphatidylcholine (PC, $\blacktriangle$), phosphatidylserine (PS, $\circ$), phosphatidylethanolamine (PE, $\bullet$), sphingomyelin (SM, $\square$), plus cholesterol in a 1:1 molar ratio. **b**, Inhibition of haemolysis by phosphorylcholine-BSA conjugates (BSA-PC, $\circ$) or as control, by BSA ($\blacktriangle$). Both phosphorylcholine and *p*-nitrophenylphosphorylcholine inhibited perforin's lytic activity at 1 mM (data not shown) indicating that the measured inhibition was indeed due to the phosphorylcholine moiety and not a result of the linker molecule used for its cross-linking to BSA.

Methods. Multilamellar liposomes of phospholipids and cholesterol were prepared by mixing 0.5 μmol of the phospholipid (Sigma) with 0.5 μmol of cholesterol in chloroform. The mixture was evaporated under reduced pressure to form a lipid film on the wall of a conical-bottomed flask. The dried lipid film was hydrated by vortexing at 50 $^{\circ}\text{C}$ in 0.1 ml of Tris-buffered saline (TBS, 10 mM Tris-HCl, pH 7.4, 150 mM NaCl). The liposome preparation was washed twice in the same buffer before use. Phosphorylcholine-conjugated BSA was prepared by incubating BSA (Fluka) with a 60-fold molar excess of *p*-diazonium phenylphosphorylcholine, synthesized as described²⁴, for 2 h at 20 $^{\circ}\text{C}$ in 0.035 M sodium borate buffer pH 9.2, containing 0.08 M NaCl. The excess reagent was removed by dialysis. Approximately 29 haptenic groups were bound per BSA molecule. The haemolytic activity of perforin was assessed as described⁷, for 2 h at 20 $^{\circ}\text{C}$ in 0.035 M sodium borate buffer pH 9.2, containing 0.08 M NaCl. The excess reagent was removed by dialysis. Approximately 29 haptenic groups were bound per BSA molecule. The haemolytic activity of perforin was assessed as described⁷. Briefly, perforin (100 μl of a concentration that had been adjusted to lyse 80% of the sheep red blood cells) was mixed with 10 μl of liposomes or BSA and immediately added to sheep erythrocytes (1.5×10^7 cells ml^{-1}) in 300 μl veronal buffered saline (GVB, 10 mM veronal buffer pH 7.4, 142 mM NaCl, 0.1% gelatin, and 0.15 mM Mg^{2+}). The mixture was incubated for 15 min at 37 $^{\circ}\text{C}$, the erythrocytes spun down and the degree of haemolysis measured by determining the haemoglobin release at 405 nm.

cytolysis. It has been shown that vesicles or lipoproteins compete with the red blood cell membrane for the binding of perforin and, hence, reduce the activity of perforin in a haemolytic assay¹³⁻¹⁵. Figure 1a shows that only vesicles containing lipids with phosphorylcholine headgroups such as sphingomyelin and phosphatidylcholine were inhibitory. This is consistent with the observation that perforin-mediated haemolysis is inhibited by 1 mM choline phosphate, whereas the presence of 50 mM ethanolamine phosphate, has no effect¹⁵. Moreover, phosphorylcholine-modified bovine serum albumin (BSA) was a potent inhibitor at 3 $\mu\text{g ml}^{-1}$ (corresponding to approximately 1.3 μM phosphorylcholine molecules), whereas no competition of perforin-induced haemolysis was observed by non-modified BSA (Fig. 1b).

Direct evidence for perforin binding to phosphorylcholine was obtained by applying perforin or a CTL-granule extract (containing perforin and the granzyme lymphocyte proteases as major components) onto a nitrophenylphosphorylcholine-Sepharose column. At normal ionic strength and in the presence of 5 mM Ca^{2+} , perforin bound to the column (Fig. 2). At Ca^{2+} concentrations below 0.1 mM, perforin was not retained (Table 1). Once perforin had bound to the column, the removal of Ca^{2+} by washing the column with ' Ca^{2+} -free' buffer did not release perforin; it only eluted after the addition of 5 mM EDTA. Ca^{2+} could not be replaced by Mg^{2+} (Table 1). Binding of perforin

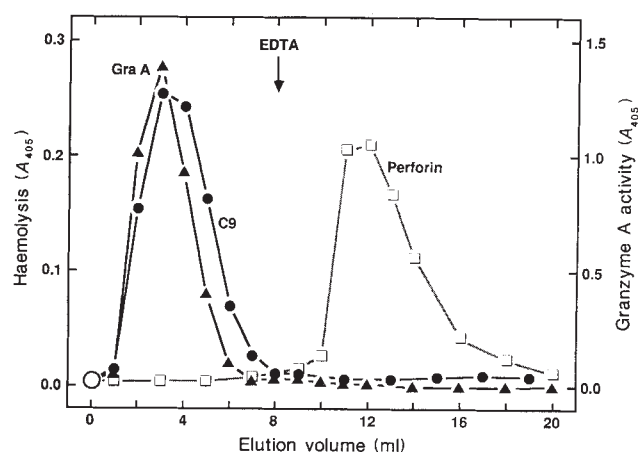


Fig. 2 Binding of perforin to a phosphorylcholine-Sepharose column. Purified perforin, an extract of granule proteins (perforin and granzyme A activity were determined) or C9 were passed over a phosphorylcholine-Sepharose column.

Methods. Phosphorylcholine affinity resin was prepared by reacting p-diazonium phenylphosphorylcholine with glycytyrosine-conjugated Sepharose 4B as described previously²⁴. Perforin and granule extracts were isolated from granules of the CTL line B6.1 (ref. 25). Granules were routinely isolated from 5×10^8 B6.1 cells. Cells were broken by nitrogen cavitation, and the organelles were separated on a discontinuous Percoll gradient. Soluble granule extracts were prepared by adding 1.5 M NaCl, thereby disrupting the granules²⁵. Perforin or granule extracts were dialysed overnight at 4 °C against TBS containing no calcium. 100 μ l aliquots were passed over a 2 ml phosphorylcholine-Sepharose column equilibrated in TBS containing 5 mM CaCl_2 . Perforin was eluted with TBS containing 10 mM EDTA. Perforin activity was determined as described above. The esterolytic activity of granzyme A was assayed using benzoyl-lysine-thiobenzylester⁵. C9 activity was determined using C5b-8-bearing erythrocytes⁷.

to phosphorylcholine was also highly dependent on the pH, as perforin was retained only above pH 6. This is consistent with the pH dependence of perforin's lytic activity (Table 1). Neither other granule proteins, that is, granzymes, nor complement C9 bound to the phosphorylcholine-Sepharose (Fig. 2).

Perforin, like C9¹⁶, has been shown to interact with membranes at 0 °C without consequent membrane insertion and lysis¹³. Perforin (or granule proteins) were incubated with red blood cells at 0 °C in the presence of 5 mM Ca^{2+} (Fig. 3). Erythrocytes were then washed twice in Ca^{2+} -containing ice-cold buffer to remove excess perforin and the subsequent increase in temperature to 37 °C led to haemolysis. In contrast, if perforin-

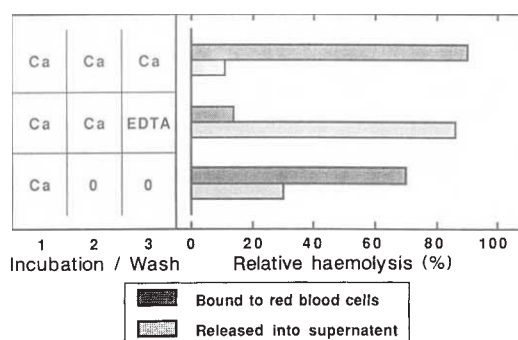


Fig. 3 Binding of perforin to red blood cells. Perforin was incubated in the presence of 5 mM Ca^{2+} with sheep red blood cells and washed and incubated: twice with a buffer containing 5 mM Ca^{2+} ; once with 5 mM Ca^{2+} and once with 10 mM EDTA; or twice with 'Ca-free' buffer. The portion of perforin bound to the red blood cells (dark bars) or released into the supernatant (light bars) was determined after the last wash/incubation step.

Methods. 400 μ l sheep red blood cells (1.5×10^7 cells ml^{-1}) in GVB containing 5 mM Ca^{2+} were incubated at 0 °C (on ice) during 10 min with a dilution of perforin or granule extracts which was three times higher than that required for 100% lysis at 37 °C after 15 min. No lysis occurred at this point. The cell suspensions were centrifuged and resuspended at 0 °C: twice in GVB containing 5 mM Ca^{2+} ; once in GVB containing 5 mM Ca^{2+} and once in GVB containing 10 mM EDTA; and twice in GVB without Ca^{2+} (' Ca^{2+} -free' buffer). This last incubation was for 10 min on ice before the red blood cells were spun down. The pellet of red blood cells was then resuspended in 400 μ l fresh GVB containing 5 mM Ca^{2+} and the mixture was incubated at 37 °C for 15 min. Perforin-induced haemolysis was determined as described in the Fig. 1 legend. The haemolytic activity of the unbound/released perforin was assayed by adding the supernatant after the last centrifugation step (300 μ l) to 100 μ l of concentrated sheep red blood cells (4.5×10^7 cells ml^{-1}) in GVB containing 5 mM Ca^{2+} (or 20 mM Ca^{2+} in the case when EDTA was used in the last wash). Haemolysis was assayed as above. The sum of lysis was arbitrarily considered as 100%.

bearing erythrocytes were washed in the presence of 10 mM EDTA, haemolysis did not occur. Perforin dissociated from the perforin-bearing erythrocyte membrane, as its activity was found in the supernatant after pelleting the intact erythrocytes. When erythrocytes were transferred into ' Ca^{2+} -free buffer', haemolysis still occurred.

These results indicate that perforin, upon exocytosis and in the presence of Ca^{2+} , binds to phosphorylcholine moieties of the target-cell membrane before lipid insertion and target-cell injury. The identification of the receptor is based on the similarity of the binding characteristics of perforin to phosphorylcholine-Sepharose and to red blood cells, respectively (Table 1). Only the binding, but not the lysis (insertion) step is dependent on the high Ca^{2+} concentrations required for perforin-mediated cytotoxicity^{1,2,4,13}. It is possible that, within the granules, perforin may have endogenously-bound ions (as has been shown for C9¹⁷) which, upon exocytosis, allow the binding of perforin to the target cells. This may partly explain recent results showing that *in vivo* primed CTL require far less Ca^{2+} than purified perforin for their cytotoxic activity^{18,19}, and that EDTA was unable to completely inhibit CTL-mediated cytotoxicity, when highly sensitive target cells were used^{20,21}.

It has been proposed that perforin and C9 have arisen from a common ancestral gene^{7,8}. A lytic protein whose receptor molecule is almost universally distributed on every cell type is potentially very dangerous for the organism. Thus, perforin is stored sequestered in cytoplasmic granules and is only released upon target-cell recognition at the site of close contact between CTL and target cell. In contrast, C9, which circulates in plasma, has lost its capacity to interact with phosphorylcholine and has become dependent on the nascent C5b-8 receptor complex for the proper target-cell recognition. The recent primary structure determination of perforin has shown that no overall homology

Table 1 Characteristics of perforin binding to phosphorylcholine-Sepharose and sheep red blood cells

Characteristics	Half-maximal binding/lysis		
	PC-Sepharose binding	Red blood cells binding	Red blood cells lysis
pH	6.5	6.5	6.5 (refs 13, 14)
[Ca^{2+}]	0.5 mM	0.5 mM	0.5 mM (refs 1, 4, 13)
[Ca^{2+}]*	<1 μ M	NA	<1 μ M
Mg^{2+} †	neg.	neg.	neg.

Experiments to investigate perforin binding to the phosphorylcholine (PC)-Sepharose or to the red blood cells, respectively were carried out as described in the legends to Figs 2 and 3, respectively, except that the pH or the metal ion conditions for the column or incubation buffer were varied. Bis-tris was used instead of Tris-HCl, when the pH was below 6.5. Concentration <1 μ M stands for ' Ca^{2+} -free' buffer. NA, not applicable; neg., negligible.

* After binding has occurred.

† 5 mM.

exists between perforin and C9^{22,23}. The structural differences at the N and C termini may account for the functional dissimilarity.

We thank E. Burnier, O. Krähenbühl, R. Etges and Z. Freiwald for their help in preparing the manuscript. J.T. is supported by a grant of the Swiss NSF.

Received 29 July; accepted 30 November 1988.

- Henkart, P. A. *A. Rev. Immun.* **3**, 31–58 (1985).
- Young, J. D.-E. & Cohn, Z. A. *Adv. Immun.* **41**, 269–332 (1987).
- Tschopp, J. & Jongeneel, C. V. *Biochemistry* **27**, 2641–2646 (1988).
- Podack, E. R. *J. Cell Biochem.* **30**, 133–170 (1986).
- Masson, D. & Tschopp, J. *Cell* **49**, 679–685 (1987).
- Müller-Eberhard, H. J. *Immunol. Rev.* **103**, 86–98 (1988).
- Young, J. D.-E., Cohn, Z. A. & Podack, E. R. *Science* **233**, 184–190 (1986).
- Tschopp, J., Masson, D. & Stanley, K. K. *Nature* **322**, 831–834 (1986).
- Tschopp, J., Podack, E. R. & Müller-Eberhard, H. J. *Nature* **298**, 534–538 (1982).
- Dankert, J. R., Shiver, J. W. & Esser, A. F. *Biochemistry* **24**, 2754–2762 (1985).
- Blumenthal, R., Millard, P. J., Henkart, M. P., Reynolds, C. W. & Henkart, P. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5551–5555 (1984).
- Criado, M., Lindstrom, J. M., Anderson, C. G. & Dennert, G. *J. Immun.* **135**, 4245–4251 (1985).
- Young, J. D.-E., Damiano, A., DiNome, M. A., Leong, L. G. & Cohn, Z. A. *J. exp. Med.* **165**, 1371–1382 (1987).
- Tschopp, J., Masson, D. & Schäfer, S. *J. Immun.* **137**, 1950–1953 (1986).
- Yue, C. C., Reynolds, C. W. & Henkart, P. A. *Molec. Immun.* **24**, 647–653 (1987).
- Bhakti, S. & Tranum Jensen, J. *J. Immun.* **136**, 2999–3005 (1986).
- Thielens, N. M., Lohner, K. & Esser, A. F. *J. biol. Chem.* **263**, 6665–6670 (1988).
- Ostergaard, H. & Clark, W. R. *J. Immun.* **139**, 3573–3579 (1987).
- Howell, D. M. & Martz, E. *J. Immun.* **140**, 1982–1988 (1988).
- Ostergaard, H. L., Kane, K. P., Mescher, M. F. & Clark, W. R. *Nature* **330**, 71–72 (1987).
- Trenn, G., Takayama, H. & Sitkovsky, M. V. *Nature* **330**, 72–74 (1987).
- Shinkai, Y., Takio, K. & Okumura, K. *Nature* **334**, 525–527 (1988).
- Lichtenheld, M. G. *et al. Nature* **335**, 448–451 (1988).
- Chesebro, B. & Metzger, H. *Biochemistry* **11**, 766–771 (1972).
- Masson, D. & Tschopp, J. *J. biol. Chem.* **260**, 9069–9072 (1985).

A kinetic intermediate in the reaction of an antigenic peptide and I-E^k

Scheherazade Sadeh-Nasseri & Harden M. McConnell

Stauffer Laboratory of Physical Chemistry, Department of Chemistry, Stanford University, Stanford, California 94305, USA

Helper T cells are triggered by molecular complexes of antigenic peptides and cell surface glycoproteins of the MHC (gene products of the major histocompatibility complex) on antigen-presenting cells¹. There is now a lot of evidence that the complexes between isolated class II MHC molecules and selected peptides have long half-lives of approximately one day^{2,3}. The reported equilibrium binding constants between antigenic peptides and class II MHC molecules however, are only micromolar^{4,5}, suggesting that the association rate constants are very low. The only reported² association rate constant is for a chicken ovalbumin peptide (OVA323–339) binding to I-A^d, and is indeed remarkably low, about 1 litre per mole per second. Prompted by these unusual data, we have used the pigeon cytochrome-c peptide pCytC(88–104) and I-E^k reconstituted in planar lipid bilayers on glass slides to investigate further the kinetics of peptide-MHC reactions. We report the formation of two I-E^k-pCytC peptide complexes. One complex has slow apparent association and dissociation kinetics, very similar to those reported previously for the chicken ovalbumin peptide and I-A^d. The second complex forms and dissociates about a hundred times more rapidly. The short-lived complex shows peptide-MHC specificity and is a kinetic intermediate in the formation of the long-lived complex; the long-lived complex is recognized by specific T-helper cells.

The helper-T-cell hybridoma 2B4 recognizes the class II MHC molecule, I-E^k and the 17 amino-acid C-terminal peptide of pigeon cytochrome-c, pCytC(88–104)⁶. This hybridoma is triggered to produce interleukin-2(IL-2) when pCytC peptide is presented by I-E^k incorporated in planar lipid bilayers on solid supports⁷. Pigeon-CytC labelled at the N terminus with fluorescein isothiocyanate FpCytC binds to I-E^k incorporated into planar membranes. The two peptides, pCytC and FpCytC, are

Table 1 Specificity of the fast reaction between pigeon cytochrome c peptide (88–104) and I-E^k in planar bilayer membranes

Reactants	Photon counts s ⁻¹
I-E ^k + FpCytC	9,864 ± 625*
I-A ^d + FpCytC	<1,000*
Lipid + FpCytC	<1,000*
I-E ^k + α I-E ^k , then FpCytC	1,100†
I-E ^k + α I-A ^k , then FpCytC	9,860†
I-E ^k + pCytC, then FpCytC	1,300‡

*Incubation was with 80 μM FpCytC for 2 min.

† Preincubation with anti-MHC antibody was for 5 min, then washed and 80 μM FpCytC added.

‡ 80 μM pCytC was incubated with I-E^k in planar membranes for 2 min and FpCytC was then added for another 2 min, then washed with PBS.

almost equally effective in triggering the 2B4 hybridoma (unpublished results). Two protocols were used to study the kinetics of FpCytC binding to I-E^k in planar membranes. In one, the fluorescence intensity attributed to FpCytC-I-E^k complexes was measured as a function of time. A reaction between FpCytC and I-E^k was essentially complete after ~100 seconds when the initial peptide concentration was 80 μM (see Fig. 1a). The binding measurements for times less than 100 s are reproducible but are of uncertain significance because of possible mixing artefacts. To test the kinetic order of the reaction a second protocol was used with series of incubations varying the incubation time (*t*) and peptide concentration (*c*) so that the product, (*ct*) was a constant (Fig. 1b). The data are not consistent with a single-step pseudo first-order reaction that goes to completion, as this should lead to a constant amount of complex when (*ct*) is kept constant.

The rapid reaction of fluorescent peptide and I-E^k was found to be specific, as (1) fluorescent OVA peptide did not bind to I-E^k, (2) preincubation with an antibody specific for I-E^k inhibited rapid binding of FpCytC to I-E^k whereas the non-specific antibody (anti-I-A^k) did not, and (3) preincubation of unlabelled pCytC with the I-E^k-containing planar membrane inhibited rapid binding of FpCytC peptide (see Tables 1, 2). The number of I-E^k molecules per unit area in the membrane (~3 × 10¹¹ cm⁻²) corresponds to the number of I-E^k molecules used in the preparation of the I-E^k-containing vesicles, and was confirmed using fluoresceinated anti-I-E^k antibody. In the experiments reported here the number of fluorescent peptides bound at maximal binding exceeded 50% of the I-E^k in the planar membrane. In a number of other experiments the maximum fluorescent peptide binding was so low (~10%) that it precluded kinetic measurements, because of a reduced signal: background ratio.

In the cellular assay, we measured the extent of T-cell stimulation (IL-2 secretion) as a function of the number of preformed complexes. The number of preformed complexes was established by incubation time and peptide concentration. The planar membranes containing the resulting I-E^k-peptide complexes were then washed, and T cells added (Fig. 1c). Again, the

Table 2 Specificity of the kinetic intermediate complex of pigeon cytochrome c and I-E^k

Reactants	Photon counts s ⁻¹
I-E ^k + FpCytC	5,800
I-E ^k + FpCytC (2 min) then pCytC*	3,800
I-E ^k + FpCytC (1 min) then pCytC	1,600
I-E ^k FpCytC (2 min) then OVA	4,600
I-E ^k + FpCytC (1 min) then OVA	4,300

Experimental conditions as described in Fig. 1a legend.

* Unlabelled peptides were incubated for 2 min at a concentration of 100 μM.

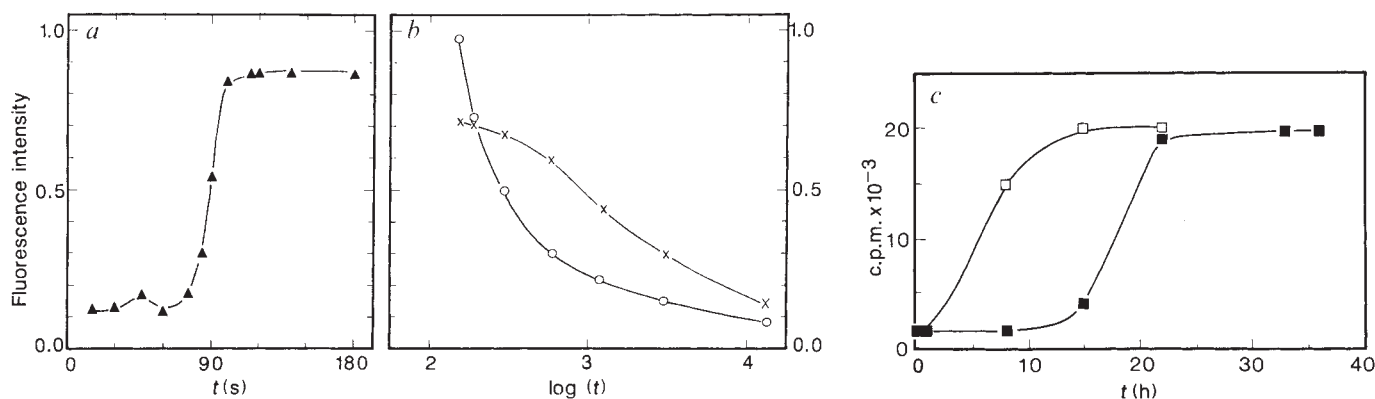
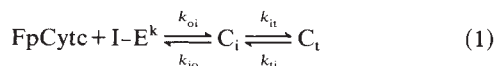


Fig. 1 Kinetics of binding of FpCytC to I-E^k in planar bilayer membranes. *a*, *b*, Fluorescence assays for the binding of FpCytC to I-E^k. *a*, Fixed 80 μM concentration of peptide versus incubation time. Fluorescence intensity is in units of 10⁴ photons s⁻¹. Data are corrected for non-specific background, which is 30–40% of total signal. *b*, Fluorescence intensity is relative to the peak intensity. ○, FpCytC at concentration *c* incubated for time *t*, such that *tc* = 12 × 10³ μM s. X indicates theoretical points based¹⁵, on estimated values of the rate constants *k*_{oi}, *k*_{ii}, *k*_{ti}, *k*_{io} given in the text, using *k*_{ii} = 4 × 10⁻⁵ s⁻¹. The values of X give the sum of the concentrations of C_i and C_t. The maximum value of X corresponds to 71% occupation of the available binding sites. A larger value of *k*_{ii} = 4 × 10⁻⁴ significantly increases (C_i + C_t) only at the two longest times. *c*, Cellular assay for the kinetics of binding of pCytC to I-E^k. Solutions of 3 μM (■) or 30 μM (□) pCytC were incubated with I-E^k in planar membranes for the indicated times, the membranes washed, 2B4 T cells added, and IL-2 secretion was assayed by [³H]thymidine uptake of the CTLL-2 cell line. The ratio of the two times (5.5 and 18 h) for 50% triggering is used to estimate the equilibrium dissociation constant *K*_i for the intermediate complex. The absolute value of each half-time can be used to estimate the rate constant *k*_{ii} for the unimolecular transition of the intermediate complex to the terminal complex under the assumption that 50% of maximal T-cell triggering corresponds to 50% occupation of available I-E^k peptide binding sites. Competition experiments show that the half times in these experiments do not depend on the number of available I-E^k binding sites (unpublished results). This analysis is equivalent to drawing first-order binding curves through the mid-points of the experimental data, and assuming that the deviation of the data from these curves is due to non-linearity in T-cell responses. **Methods.** I-E^k was purified by affinity chromatography using monoclonal antibody 14-4-4S as described^{9,10}. Lipopolysaccharide (*Escherichia coli* 0.111:84, Difco) stimulated blast cells from spleens of CBA (H-2^k) mice were used as the source of I-E^k. Reconstitution of I-E^k into planar membranes was as described¹⁰. Vesicle preparations for cellular assays used the lipid mixture of dipalmitoylphosphatidylcholine and dilinoleoyl phosphatidylcholine and cholesterol⁷. In *a*, 80 μM FpCytC was incubated with planar membranes containing I-E^k for the times indicated, then washed for one minute with PBS, all at room temperature (20 °C). Fluorescence intensity was measured with the photon counting system described previously⁸. The incubations were carried out in chambers of 20 μl volume assembled on a quartz microscope slide and covered by a circular 12-mm diameter coverglass using two layers of double sticky tape. Each microscope slide had two chambers. Vesicle preparations with or without I-E^k were deposited in the chambers, allowing planar membrane formation. N-terminal fluorescence of pCytC (88–104) was carried out during solid phase peptide synthesis (Peninsula), and purified by reverse phase HPLC on a Vydac C₁₈ column. In *c*, the cell-free reaction took place during time intervals shown. Excess unbound peptide was removed by careful replacement of the medium covering the planar membrane, over a period of 10 min or more. T cells were added at a density of 4 × 10⁵ to every well. The amount of IL-2 secreted in culture wells was assayed by [³H]thymidine uptake of an IL-2-dependent T-cell line CTLL-2.

binding kinetics cannot be accounted for by a single-step pseudo first-order reaction. That is, the times for 50% cell triggering are not inversely proportional to the peptide concentration.

The experiments in Fig. 2 (▲) show the dissociation of the FpCytC-I-E^k complex as a function of time. FpCytC and I-E^k in planar membranes were incubated at room temperature for various lengths of time and were then sequentially rinsed to remove dissociated FpCytC. The wash was repeated at the indicated times for 2.5 days. After many washes the samples lose only 5–10% of the I-E^k as judged by the binding of anti-I-E^k. The dissociation curve is evidently biphasic with fast and slow components. Both reactions follow first-order kinetics with half lives of approximately 10 minutes and of 30 hours. The half life for the dissociation of pCytC from I-E^k at 37 °C obtained from the cellular assay is similar: 34 hours (Fig. 2, ●).

The dissociation data can be understood semiquantitatively in terms of two complexes of FpCytC and I-E^k, a short-lived intermediate complex C_i and a long-lived terminal complex C_t.



The first-order dissociation rate constant for the intermediate of FpCytC complex in Fig. 2a is *k*_{io} = 1.2 × 10⁻³ s⁻¹. The first-order dissociation rate constant for the long-lived terminal complex is *k*_{io} ≈ *k*_{ii} = 6.3 × 10⁻⁶ s⁻¹. The relative amplitudes of the slow and fast components of dissociation depend on the time of preincubation before the dissociation measurement. The longer incubation time gives a larger proportion of the slowly dissociating component. Thus a 5-min incubation gives ~15%

of the slow component, whereas a 20-min incubation gives ~30% of the slow component, both with 80 μM FpCytC.

These results can be interpreted using a first-order rate constant for the conversion C_i to C_t of *k*_{ti} = 4 × 10⁻⁴ s⁻¹. The data in both Fig. 1c and Fig. 2 provide evidence that C_i is an intermediate in the formation of C_t and is not some irrelevant dead-end complex. The experiments do not show whether C_i is a complex with only one bound peptide (FpCytC) or two peptides (for example, FpCytC and a second peptide). The slowly decaying component C_t can be identified as biologically active, given the excellent agreement of the low off-rates based on the fluorescence and cellular assays (Fig. 2, ●).

According to reaction (1) the resultant kinetics of the formation of the terminal complex C_t from peptide and I-E^k are pseudo first-order with a rate constant that is non-linear in the peptide concentration, *k*_{ot} = (*k*_{ti})/(*p* + *K*_i), where *p* is the peptide concentration and *K*_i is the dissociation constant of the intermediate complex C_i. The data in Fig. 1c can be interpreted with *K*_i ≈ 10⁻⁵ M and *k*_{ti} ≈ 4 × 10⁻⁵ s⁻¹; this value for *k*_{ti} assumes the cellular response is half-maximal at 50% saturation of the I-E^k with pCytC. The disparity between the two estimates of *k*_{ti} is not unreasonable in view of the uncertainties in this assumption. We estimate the equilibrium dissociation constant *K*_t of the terminal complex from the equation *K*_t = (*k*_{ii}/*k*_{ti})*K*_i and obtain *K*_t ~ 10⁻⁶–10⁻⁷ M, depending on the choice of *k*_{ti} from the above estimates. Complexes of other antigenic peptides and Ia studied by Babbitt and co-workers⁴ and Buus and co-workers⁵ have dissociation constants *K*_t ~ 10⁻⁶ M. Our model reduces mathematically to that of Buus at low peptide concentrations (*p* < *K*_i) where only three constants are observed, *K*_t, *k*_{oi} and *k*_{io}.

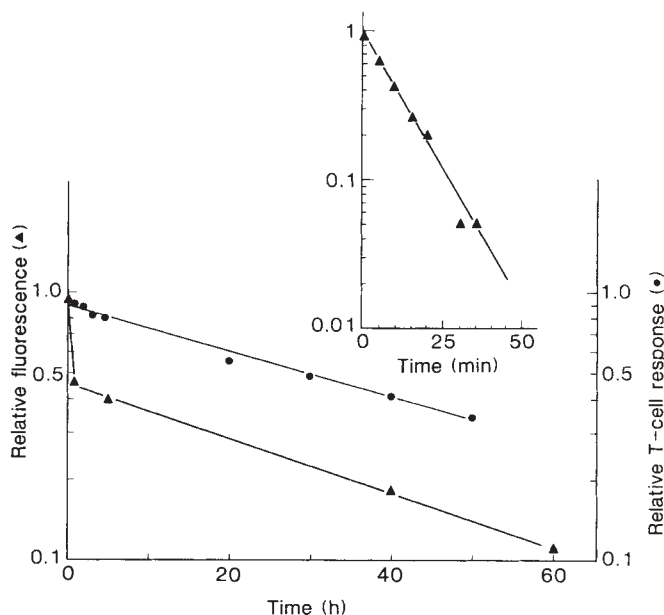


Fig. 2 Kinetics of dissociation of pigeon cytochrome-c peptide (88-104)-I-E^k complex. Fluorescence assay, ▲, curve and insert). Following an ~25 min incubation of I-E^k in planar membrane together with 80 μ M FpCytC, the sample was washed sequentially at the indicated times. Each wash lasted about 1 min, and involved flowing 10-15 volumes PBS through the sample chamber. The data given in the insert were obtained after a 5-min incubation. Fluorescence intensities in the insert are used to estimate the off-rate constant k_{io} for the intermediate complex, and are corrected for both background fluorescence and the fluorescence intensity of the long-lived complex. After this 5-min incubation the short lived fluorescence (10-min half-life) is about 85% of the total specific fluorescence. The half-time for the dissociation of FpCytC from the terminal complex from the fluorescence data is 30 hours. In the cellular assay, (●) 15-20 μ M unlabelled peptide was incubated at 37 °C with planar membranes with I-E^k for 24 h in RPMI medium without serum. Excess unbound peptide was removed. Each data point was obtained by allowing dissociation of peptide from I-E^k-peptide complexes followed by a second rinse before addition of 2B4 T cells in 10% fetal calf serum. The half time obtained from these cell assay data is 34 h. At each time point, the proliferation of CTLL was measured by titrating the supernatant and using data in the linear range.

The kinetic scheme (1) does not account for the observed but uncertain shape of the short time (<100 s) binding curve. A value of $k_{oi} = 1.2 \times 10^2 \text{ l mol}^{-1} \text{ s}^{-1}$ obtained from the above estimates of K_i and k_{io} leads to a half-time for the formation of C_i of 70 seconds at 80 μ M peptide, a result close to the onset time of ~90 seconds seen in Fig. 1a. These estimated values of K_i and k_{oi} also lead to the calculated points X in Fig. 1b, to be compared with the experimental points O. The calculations used the solution for two sequential first-order reactions given by Moore and Pearson¹⁵. At short times and high peptide concentrations the fraction of the available I-E^k binding sites occupied by the peptide is approximately $pk_{oi}/k_{it}t^2$, so that even at times as short as 100 s as much as 1% terminal complex can form, which may account for some of the very recent observations on the kinetics of peptide-MHC complex formation, as measured by calcium influx in specific T cells¹⁶.

In summary, a variety of kinetic data all point to the formation of an intermediate complex between an antigenic peptide and an MHC molecule. This result is particularly significant as the intermediate complex shows peptide-MHC specificity. The molecular structures of the intermediate and terminal complexes may be very different as the time required for the unimolecular conversion of the intermediate complex to the terminal complex is long, estimated to lie between 30 and 300 min. The slow step may or may not involve the release of a second peptide. Thus,

although specific helper T cells presumably recognize the structure of the terminal complex, the selective peptide recognition by MHC may involve a substantially distinct protein-peptide conformation¹²⁻¹⁴.

We thank Dr Gary Fathman for the use of his cell harvester, and Dr Benno Rothenhäusler for bringing ref. 15 to our attention. This work was supported by the NIH. S.S.N. is a recipient of Cancer Research Institute Fellowship award.

Received 20 July; accepted 30 November 1988.

1. Schwartz, R. H. *A. Rev. Immun.* **3**, 237-261 (1985).
2. Buus, S., Sette, A., Colon, S. M., Jenis, D. N. & Grey, H. M. *Cell* **47**, 1071-1077 (1986).
3. Watts, T. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9660-9664 (1986).
4. Babbitt, B., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
5. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-3971 (1986).
6. Samuelson, L. E., Germain, R. N. & Schwartz, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972-6976 (1983).
7. Quill, H., Carlson, L., Fox, B. S., Weinstein, J. N. & Schwartz, R. M. *J. Immunol. Meth.* **98**, 29-41 (1987).
8. Watts, T. H., Gaub, H. E. & McConnell, H. M. *Nature* **320**, 179-181 (1986).
9. Ozata, K., Mayer, N. & Sachs, D. H. *J. Immun.* **124**, 533-540 (1980).
10. Watts, T. H., Brian, A. A., Kappler, J. W., Marrack, P. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7564-7568 (1984).
11. Gillis, S. & Smith, K. A. *Nature* **268**, 154-156 (1977).
12. Bjorkman, P. J. *et al. Nature* **329**, 506-512 (1987).
13. Bjorkman, P. J. *et al. Nature* **329**, 512-518 (1987).
14. Brown, J. H. *et al. Nature* **332**, 845-850 (1988).
15. Moore, J. W. & Pearson, R. G. *Kinetics and Mechanism*, 296-300 (Wiley, New York, 1981).
16. Roosnek, E., Demetz, S., Corradin, G. & Lanzavecchia, A. *J. Immun.* **140**, 4079-4082 (1988).

Identification of specific binding proteins for a nuclear location sequence

Stephen A. Adam*, Thomas J. Lobl†‡, Mark A. Mitchell† & Larry Gerace*

* Department of Molecular Biology, Research Institute of Scripps Clinic, 10666 N. Torrey Pines Road, La Jolla, California 92037, USA
† The Upjohn Company, Kalamazoo, Michigan 49001, USA

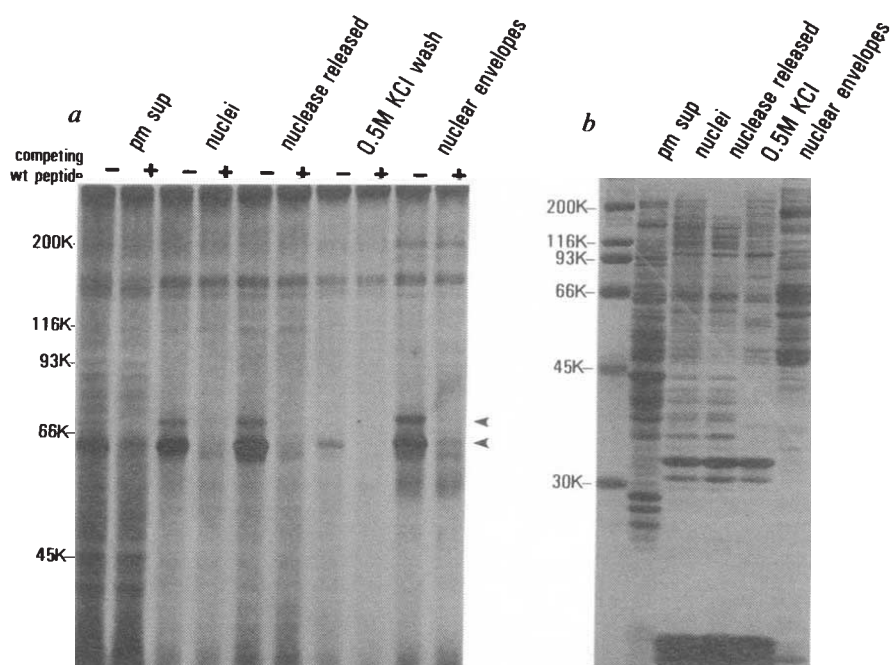
The nuclear envelope is a selective barrier against the movement of macromolecules between the nucleus and cytoplasm¹. Nuclear proteins larger than relative molecular mass 20,000-40,000 are probably actively transported across the envelope through the nuclear pore complex^{2,3} and are directed by specific nuclear location sequences (NLS) in the proteins⁴. NLS mediate the nuclear import of isolated nuclear proteins after microinjection into whole cells^{5,6} and the nuclear accumulation of chimaeric proteins⁷⁻⁹ or of non-nuclear proteins conjugated to synthetic peptides¹⁰⁻¹². The best-characterized NLS is the simian virus 40 large T-antigen sequence^{4,13,14}. We have identified two proteins of rat liver by chemical cross-linking that interact with a synthetic peptide containing this sequence: this interaction is specific for a functional NLS, is saturatable, and high affinity. The binding proteins are present in a post-mitochondrial supernatant, in nuclei and in a nuclear envelope fraction, which is consistent with a role in the transport of nuclear proteins from the cytoplasm to the nucleus.

Chemical cross-linking with disuccinimidyl suberate of a radioiodinated synthetic 34-residue peptide from the SV40 large T antigen containing the NLS was used to identify the proteins in different subcellular fractions of rat liver that interact with the NLS (Fig. 1). A 10-residue synthetic peptide containing the wild-type NLS sequence was used as a competitor to define binding specificity. Two proteins that interact specifically with the NLS, and can be competed out by a 100-fold molar excess of the 10-residue wild-type peptide, are detected in a post-

‡ Present address: Immunotech Pharmaceuticals, 11045 Roselle Street, San Diego, California 92121, USA

Fig. 1 Intracellular distribution and specificity of the SV40 large T antigen nuclear location sequence binding proteins. **a**, Equivalent amounts of protein from octylglucopyranoside/high salt extracts of post-mitochondrial supernatant (pm sup, 0.05 nuclear equivalents (n.e.) where 1 n.e. = 3×10^6 nuclei), whole nuclei (0.68 n.e.), nuclease-released soluble (1 n.e.), 0.5M KCl-wash (8.7 n.e.) and nuclear envelopes (14 n.e.) were incubated with radiolabelled peptide and cross-linked with disuccinimidyl suberate (Pierce). An autoradiograph of the sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel is shown. Arrows indicate specific bands. **b**, Coomassie-blue-stained gel of the subcellular fractions used in **a**. The lower portion of the figure presents the sequences of peptides used for cross-linking and competition. The underlined residues represent the minimum SV40 large T antigen nuclear location sequence. The 34-residue peptide was labelled with ^{125}I . Molecular weights of marker proteins are indicated.

Methods. Rat liver nuclear envelopes were prepared as described¹⁷. Post-mitochondrial supernatant was prepared by centrifugation of the first post-nuclear supernatant at 12,500g for 15 min. Portions of each fraction containing 50 μg protein were incubated in 20 μl buffer A (15 mM HEPES-KOH, pH 7.3, 0.25 mM MgCl_2 , 8.5% sucrose, 1 $\mu\text{g ml}^{-1}$ each leupeptin and pepstatin A, 2 mM dithiothreitol, 10 u ml^{-1} aprotinin) 1% octyl- β -D-glucopyranoside (Calbiochem) and 300 mM KCl for 1 hour on ice. The extracts were diluted sixfold in buffer A and 50 nM ^{125}I -labelled SV40 large T antigen peptide or the 10-residue peptides was added to 5 μM for competition. The 34-residue analogue was labelled with ^{125}I using Iodogen (Pierce) to a specific activity of $3\text{--}5 \times 10^5$ c.p.m. ng^{-1} peptide. After 30 min at 25 $^\circ\text{C}$, the samples were cooled on ice for 5 min and disuccinimidyl suberate was added to 0.5 mM from a 50 mM stock in dimethylsulphoxide. Cross-linking was terminated after 15 minutes by precipitation with trichloroacetic acid added to 10%; precipitates were collected by centrifugation, washed with 95% acetone and air-dried before resuspension in SDS-PAGE sample buffer. Samples were resolved in 8% SDS-PAGE gels¹⁸. The gel in **b** is 11% polyacrylamide, rather than 8%, which resolves the histones in the nuclear fractions. The gels were fixed and stained with Coomassie blue, destained and dried before autoradiography using Kodak XAR-5 film with an intensifying screen at -80°C .



Labelled peptide

cys-tyr-asn-asn-glu-ala-thr-ala-asn-ser-gln-his-
ser-thr-pro-pro-lys-lys-lys-arg-lys-val-glu-
asn-pro-lys-asn-phe-glu-ser-glu-leu-leu-ser-NH₂

Competing peptide

Wild type cys-pro-lys-lys-lys-arg-lys-val-glu-asn-NH₂

mitochondrial supernatant, in nuclei and in a highly purified nuclear envelope fraction by the cross-linking procedure (see arrows in Fig. 1a and b); these are a predominant protein of relative molecular mass (M_r) 60,000 (60K) and a minor protein of 70K (these values include the mass of the cross-linked peptide). The 70K protein in the post-mitochondrial supernatant can be difficult to visualize (as in Fig. 1), but usually is evident (data not shown). Isolation of nuclear envelopes is by nuclease digestion of nuclei to give a nuclease-released fraction (Fig. 1), followed by extraction in 0.5M KCl to solubilize residual chromatin (0.5M KCl wash, Fig. 1). Most of the NLS-binding proteins are released from nuclei by nuclease digestion, which is consistent with a partially intranuclear location. In different experiments, however, we find that 10–30% of the nuclear-binding activity is retained in the nuclear envelope fraction in a salt-resistant fashion. This indicates that at least a fraction of the NLS-binding proteins are associated with the nuclear envelope. The activity detected in the post-mitochondrial supernatant fraction is probably not a result of release of the proteins from the nucleus during preparation, because isolated nuclei incubated at 4 $^\circ\text{C}$ for several hours in the homogenization buffer do not release the binding proteins (data not shown). The binding proteins in the post-mitochondrial supernatant are retained in a high-speed supernatant and therefore are not

associated with any membranes (data not shown). Although these data suggest that the location for NLS-binding proteins is partially cytoplasmic, definitive localization requires antibodies specific to these proteins for immunocytochemistry.

To detect the NLS-binding proteins optimally in nuclear envelopes, it was necessary to extract them with 1% octyl- β -D-glucopyranoside/0.3M KCl before cross-linking (Fig. 2). Substantially more cross-linking of the peptide to the NLS-binding proteins was seen in the detergent-soluble fraction (Fig. 2, OCG/KCl in lane S) compared with intact nuclear envelopes (buffer), with essentially no detectable binding protein remaining in the insoluble lamina fraction (OCG/KCl, lane P). The reason for this binding 'latency' is unclear. As these detergent and salt conditions completely solubilize some nuclear pore complex proteins¹⁵ (data not shown), it is possible that the masked binding activity represents NLS-binding proteins associated with endogenous transport ligands which become dissociated in the presence of detergent and high salt. In contrast to nuclear envelopes, treatment of the other subcellular fractions with detergent and salt resulted in only a small increase in the binding proteins detected (data not shown).

The 60K and 70K labelled proteins were present in the same relative abundance in all subcellular fractions, but their relationship is unknown. The association of the synthetic peptides with

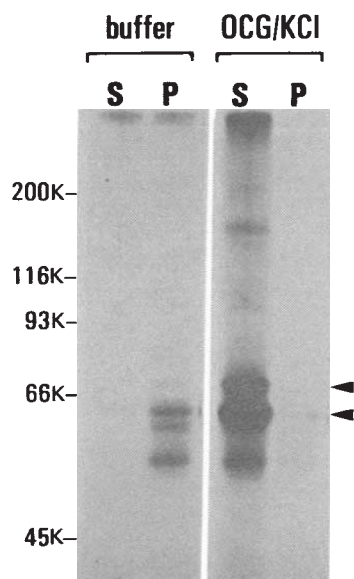


Fig. 2 Extraction of SV40 nuclear location sequence binding protein from nuclear envelopes. The binding proteins are masked in the nuclear envelope (buffer) but can be exposed by extraction in 1% octyl- β -D-glucopyranoside/0.3M KCl (OCG/KCl). Detergent-soluble (S) and insoluble material (P) after extraction and cross-linking are shown on an autoradiograph of the SDS-PAGE gel.

Methods. 50 μ g nuclear envelopes were incubated in buffer A containing 50 mM KCl (buffer) or 1% OCG/300 mM KCl (OCG/KCl) as described in Fig. 1 legend. Soluble and insoluble fractions were separated after cross-linking by centrifugation at 5,000g for 15 min.

the binding proteins occurs rapidly, reaching equilibrium within minutes at 25°C, and does not require ATP (data not shown). The binding proteins associated with the solubilized nuclear envelope (Fig. 2) are not immunoadsorbed with monoclonal antibodies to nuclear pore complex glycoproteins¹⁵ and are therefore distinct from them (data not shown).

We used the peptide cross-linking assay to test the concentration dependence of the interaction of the 60K protein with the 34-residue peptide (Fig. 3). Saturation was achieved in all fractions at 50–75 nM peptide concentrations, which indicated a high-affinity interaction. The binding curves for the post-mitochondrial supernatant (Fig. 3a) and the nuclear envelope (Fig. 3c) material are hyperbolic, reaching half-maximal saturation at 30 nM peptide, whereas the nuclease-released (Fig. 3b) material gives a sigmoidal curve, with half-maximal saturation at almost 60 nM peptide. Whether this represents a true difference in binding properties or results from other differences between the fractions is unknown.

The specificity of binding was further analysed by competition with peptides containing sequences previously characterized by their ability to direct nuclear transport *in vivo*^{14,16}. A 10-fold and 100-fold molar excess of each peptide was used to compete with specific binding of the 34-residue NLS peptide (Fig. 1c) in the cross-linking assay. A peptide containing the wild-type NLS (Fig. 4, wild type), as well as a peptide containing a mutation that does not affect transport *in vivo* (peptide A) strongly compete for binding with the NLS peptide. A peptide containing a mutation that abolishes transport *in vivo* (peptide B) does not inhibit binding, whereas two peptides containing sequences that allow partial transport compete weakly or not at all (peptides C and D). The observed competition with peptides containing transport-competent sequences is not a simple consequence of charge density on the peptides, as a peptide containing the reverse of the wild-type sequence (that is, from

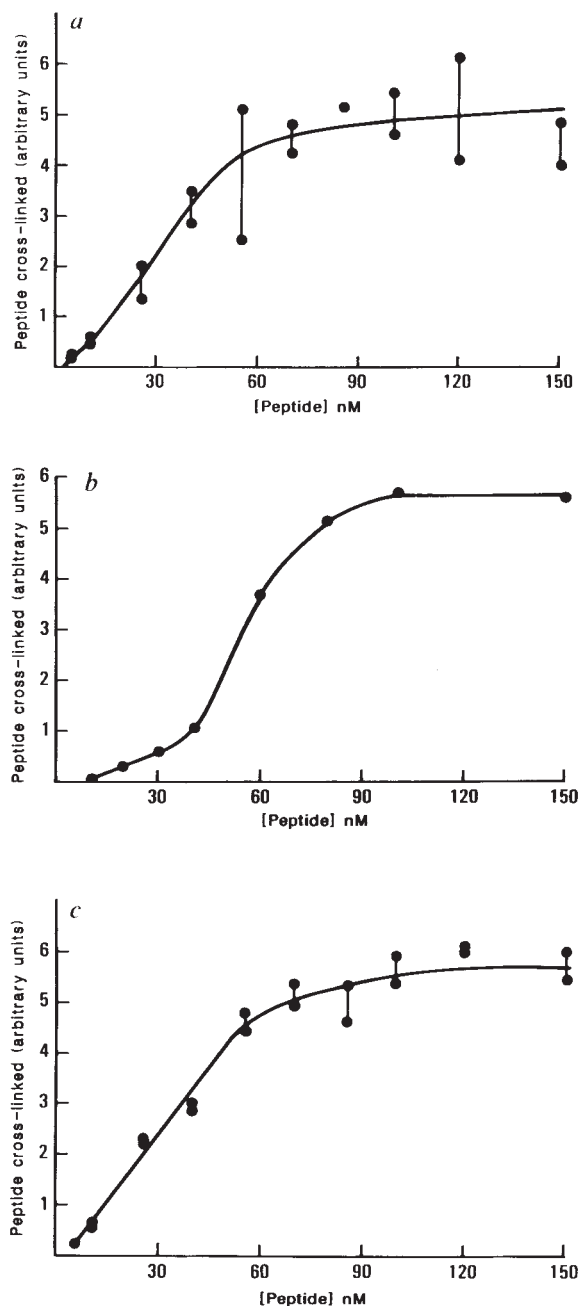


Fig. 3 Saturatable, high-affinity binding of the NLS peptide to the NLS-binding proteins. Cross-linking to (a) post-mitochondrial supernatant, (b) nuclease-released material or (c) nuclear envelopes demonstrates that the binding is saturatable and of high affinity. Data represent quantitation of the 60K protein; equal amounts of protein were analysed for each fraction.

Methods. Subcellular fractionation and cross-linking are as described in Fig. 1. Autoradiographs were scanned in an LKB Ultrascan XL laser densitometer. Peak areas were integrated automatically and used to plot the relative amount of radioactivity in each band. Arbitrary units were assigned to the nuclear envelope fraction and the values for the other fractions were adjusted to give similar values at saturation. Samples were run in duplicate, with the exception of the nuclease-released fraction which was determined once at each peptide concentration. Cutting individual bands from the gels and counting in a gamma counter gave similar results.

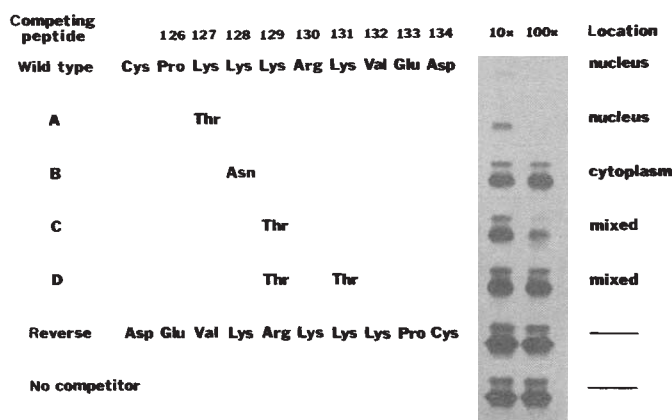


Fig. 4 Specific binding of the NLS peptide to the NLS-binding proteins. Cross-linking was as described in Fig. 1, using binding proteins extracted from nuclear envelopes. Cross-linking in the presence of 10-fold and 100-fold molar excess of competing peptides is shown. Subcellular locations were taken from ref. 14.

Methods. Wild-type, B and reverse peptides, as well as the 34-residue peptide, were synthesized by the stepwise coupling of activated protected amino acids by the solid-phase peptide synthesis method¹⁹ on to 4-methylbenzhydrylamine resin using an Applied Biosystems 430A automated peptide synthesizer. The peptides were cleaved from the resin by liquid HF at 0°C for 1 hour in the presence of anisole¹⁹. The 34-residue analogue was cleaved by the low/high HF method²⁰, modified to include cysteine-HCl. The crude peptides were purified by HPLC on Vydac RP-C18 or RP-C4 preparative columns using a water/acetonitrile-containing 0.1% trifluoroacetic acid (gradient 100/0 to 50/50). Peptides A, C and D were synthesized by the simultaneous multiple-peptide synthesis method²¹ and were not further purified for this analysis. All peptides were further purified by chromatography on Sephadex G-10 in 0.1M acetic acid and concentrations of each were determined by Ellman's reaction²².

carboxyl to amino residues) does not compete for binding (peptide reverse). Thus there is a close correspondence between the ability of different sequences to direct transport *in vivo* and their ability to compete with NLS-peptide binding to the 60K and 70K proteins *in vitro*.

These cross-linking experiments identify only two proteins in rat liver that specifically interact with the SV40 large T antigen nuclear location sequence. As the interaction is also high-affinity and saturable, it is most likely that the NLS-binding proteins we have detected represent carrier molecules for nuclear import of proteins containing a functional SV40 large T-antigen NLS. The apparent distribution of these proteins in the cytoplasm and nuclear interior as well as in the nuclear envelope supports a multistep model for carrier-mediated nuclear protein import. The initial step, binding of the protein to be transported to a receptor in the cytoplasm, could occur very soon after translation; this receptor-protein complex could then be recognized at the nuclear pore complex by another receptor molecule and be translocated through the pore channel, concomitantly with channel gating³. The internalized complex could then dissociate, with the free receptor being recycled to the cytoplasm or degraded. Further characterization of the NLS-binding proteins described here in conjunction with *in vitro* and whole-cell nuclear transport systems will address some of these possibilities.

We thank Dr Richard Houghten and Linda L. Maggiora for synthesis of some of the peptides, and Drs J. Glass and C. Featherstone for discussion and comments on the manuscript. S.A.A. was supported by the Damon Runyon-Walter Winchell Cancer Research Fund. This work was supported by a grant from the American Cancer Society to L.G.

Note added in proof: The reverse sequence peptide in Fig. 4 has recently been shown to be incapable of directing nuclear transport *in vivo* (T. J. Lobl, personal communication).

Received 2 November; accepted 24 November 1988.

- Gerace, L. & Burke, B. A. *Rev. Cell Biol.* **4**, 335-374 (1988).
- Feldherr, C. M., Cohen, R. J. & Ogburn, J. A. *J. Cell Biol.* **96**, 1486-1490 (1983).
- Feldherr, C. M., Kallenbach, E. & Schultz, N. J. *J. Cell Biol.* **99**, 2216-2222 (1984).
- Dingwall, C. & Laskey, R. A. *Rev. Cell Biol.* **2**, 367-390 (1986).
- Lanford, R. E., Kanda, P. & Kennedy, R. C. *Cell* **46**, 575-582 (1986).
- Goldfarb, D. S., Garipey, J., Schönik, G. & Kornberg, R. *Nature* **322**, 641-644 (1986).
- Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. *Cell* **39**, 499-509 (1984).
- Hall, M. N., Hereford, L. & Herskowitz, I. *Cell* **36**, 1057-1065 (1984).
- Burglin, T. R. & DeRobertis, E. M. *EMBO J.* **6**, 2617-2625 (1987).
- Newmeyer, D. D., Lucocq, J. M., Burglin, T. R. & DeRobertis, E. M. *EMBO J.* **5**, 501-510 (1986).
- Newmeyer, D. D., Findlay, D. R. & Forbes, D. J. *J. Cell Biol.* **103**, 2091-2102 (1986).
- Markland, W., Smith, A. E. & Roberts, B. L. *Molec. cell. Biol.* **7**, 4255-4265 (1987).
- Lanford, R. E. & Butel, J. S. *Cell* **37**, 801-813 (1984).
- Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. *Nature* **311**, 33-38 (1984).
- Snow, C. M., Senior, A. & Gerace, L. *J. Cell Biol.* **104**, 1143-1156 (1987).
- Lanford, R. E., White, R. G., Dunham, R. G. & Kanda, P. *Molec. Cell. Biol.* **8**, 2722-2729 (1988).
- Dwyer, N. & Blobel, G. *J. Cell Biol.* **70**, 581-591 (1976).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Stewart, J. M. & Young, J. D. *Solid Phase Peptide Synthesis* (Pierce Chem. Co., Rockford, Illinois, 1984).
- Tam, J. P., Heath, W. F. & Merrifield, R. B. *J. Am. chem. Soc.* **105**, 6442-6455 (1983).
- Houghten, R. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5131-5135 (1985).
- Ellman, G. L. *Arch. Biochem. Biophys.* **74**, 443-450 (1958).

Role of transcriptional interference in the *Drosophila melanogaster* *Adh* promoter switch

Victoria Corbin & Tom Maniatis

Harvard University, Department of Biochemistry and Molecular Biology, Cambridge, Massachusetts 02138, USA

The *Drosophila melanogaster* alcohol dehydrogenase (*Adh*) gene is transcribed from two closely linked promoters^{1,2}, which are regulated by two upstream enhancers. The proximal promoter is active primarily in first to early third-instar larvae, whereas the distal promoter is active in late third-instar larvae and adults (Fig. 1). The *Adh* larval enhancer and the proximal promoter are separated by the *Adh* adult enhancer and the distal promoter. Because the proximal promoter is turned off just as the distal promoter is turned on, we considered the possibility that the distal promoter or adult enhancer has a role in the downregulation of the proximal promoter. We report here that transcription from the distal promoter is required to shut off the proximal promoter. In the absence of the distal promoter, the proximal promoter is active throughout larval development and in adults. The proximal promoter is also aberrantly active in adults when placed upstream of the distal promoter. These results suggest that the developmental switch from proximal to distal promoter is regulated by the stage-specific activation of the distal promoter, and the subsequent repression of the proximal promoter by transcriptional interference³⁻⁸.

Initially we deleted the distal promoter and the *Adh* adult enhancer (AAE) and examined the activity of the proximal promoter at various developmental stages. When the distal promoter and AAE were deleted (Fig. 2) the proximal promoter was not shut off in mid third-instar larvae but, rather, remained active throughout larval development (Fig. 3a, compare lanes 4-6 with 11-14). In addition, the proximal promoter was expressed at aberrantly high levels in adults (Fig. 3a, compare lanes 7 and 15). Thus, the distal promoter and AAE are required for the normal downregulation of the proximal promoter in third-instar larvae and in adults.

A deletion that removes the TATA box and initiation site of the distal promoter (-66 to +53), also leads to the derepression of the proximal promoter in adults (Fig. 4a, lanes 12-14). This result suggests that readthrough transcription from the upstream, distal promoter inactivates the proximal promoter (transcriptional interference³⁻⁸). Unlike other systems where transcrip-

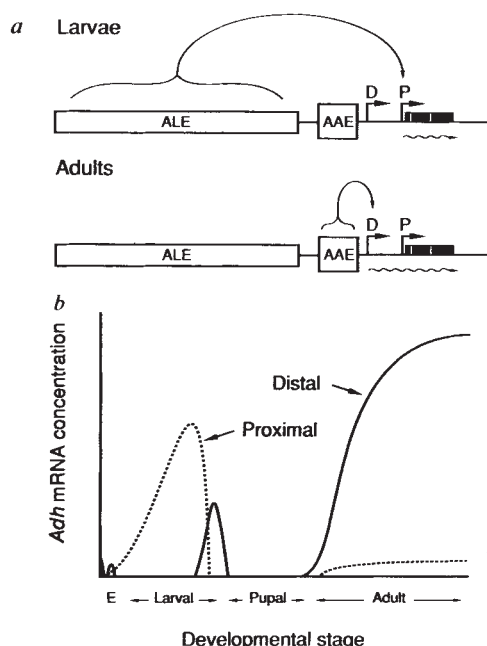


Fig. 1 The organization of *Adh* regulatory sequences and the pattern of *Adh* expression during development. *a*, The *D. melanogaster Adh* gene, showing the enhancer-promoter interactions in different developmental stages. In larvae, the *Adh* larval enhancer (ALE) stimulates transcription primarily from the proximal promoter (V.C. and T.M., manuscript in preparation). ALE activity maps to three disparate regions which lie between -5,000 and -660 base pairs (bp) of the distal promoter (V.C. and T.M., unpublished results). In adults, the *Adh* adult enhancer (AAE) stimulates transcription from the distal promoter¹³. The AAE is located between -590 and -128 bp of the distal promoter (D. Falb and T.M., unpublished observations). *b*, The relative levels of proximal (dotted line) and distal (smooth line) promoter transcripts are plotted as a function of developmental stage (taken from ref. 2).

tional interference has been detected^{4,5}, deletions that reduced the activity of the upstream (distal) *Adh* promoter, did not proportionately increase the activity of the downstream (proximal) promoter (Fig. 4a, lanes 6-11). This observation does not, however, eliminate the transcriptional interference model, as the levels of transcription from the weakened distal promoters are still relatively high and could therefore be sufficient to inactivate the proximal promoter.

To test whether the proximal promoter is repressed in adults by transcriptional interference, we placed a second copy of the proximal promoter and *Adh*-coding region upstream of the AAE and distal promoter, so that transcription from the distal promoter could not read through the upstream proximal promoter (5'P3'DP, Fig. 2). In this context, the proximal promoter was highly active in adults, even though the distal promoter was expressed at the usual high level (Fig. 4b, lanes 7-9). Thus, when transcription from the distal promoter cannot read through the proximal promoter, the proximal promoter is not shut off in adults.

These results are consistent with the transcriptional interference model, but they do not exclude the possibility that the distal promoter, by virtue of its location, blocks interactions between the proximal promoter and the *Adh* larval enhancer (ALE) or the AAE (see Fig. 1a; although the AAE does not normally stimulate the proximal promoter, it can if the distal promoter is deleted¹³). We tested this possibility by placing the ALE or AAE at the 3' end of the gene, so that the distal promoter does not lie between the enhancer and the proximal promoter. The stage-specific expression of the two promoters was not altered when the ALE was placed downstream from the gene (V.C. and T.M., manuscript in preparation). Similarly, the proximal promoter was repressed normally in adults when the AAE was placed downstream from an *Adh* gene containing both promoters. (Fig. 4b, lanes 5 and 6). In contrast, the 3'AAE did activate the proximal promoter in adults when the distal promoter was deleted (data not shown). We conclude that the stage-specific activity of the proximal promoter is independent

Fig. 2 Structure and expression of mutant *Adh* genes. The structure (not to scale) and transcriptional activity of *Adh* gene constructs in adult transformants are shown. In wild-type adults, the distal promoter is active at high levels (+) and the proximal promoter is active at low levels (-). Different levels of transcripts were observed with independent transformants containing the same *Adh* mutants. This variability is presumably due to the influence of surrounding chromosomal sequences or structure¹⁴. Therefore, we have indicated the activity of a given gene as a range. N/A, not applicable. Deletion endpoints are numbered relative to the transcription start site of the distal promoter. The enhancers (boxes), promoters (D and P) are indicated. The D-5000 gene includes sequences from -5,000 to +2,510 and is expressed at wild-type levels in all developmental stages (V.C. and T.M. manuscript in preparation). The ALE/P-386 gene contains the ALE fused directly to -386 of the proximal promoter, deleting the AAE and the distal promoter (-386 of the proximal promoter equals +328 of the distal promoter). ALE/P-386 transformants produce proximal transcripts at levels ~3-fold (1 line), 5- to 10-fold (2 lines), 15- to 20-fold (2 lines), and ~30-fold (1 line) higher than wild-type controls. The internally deleted genes, D-660/-128, D-127/D-63, D-66/+53 and D+53/+328, remove the ALE (sequences between -660 and -128), sequences just upstream of the distal promoter (between -127 and -63), the distal promoter's TATA box and initiation site (sequences between -66 and +53), and the sequences downstream of the distal promoter (between +53 and +328). In D-66/+53 transformants, the activity of the proximal promoter was three-fold (3 lines) or 5- to 10-fold (4 lines) higher than wild-type, although no proximal promoter activity was observed in 2 other lines. D-660 contains sequences from -660 to +2,510 (V.C. and T.M. manuscript in preparation). 5'3'AAE contains one copy of the AAE in its normal location, and a second copy (-660 to -128) at the 3' end of the gene (+3,100). 5'P3'DP contains a duplicate proximal promoter plus *Adh* gene body (+328 to +2,510) linked to a downstream *Adh* gene (-660 to +3,100). P-386 contains sequences from +328 (-386 of the proximal promoter) to +2,510 of the distal promoter. **Methods.** Standard cloning procedures were used¹⁵. *Adh* constructs were cloned into the transformation vector Carnegie 20 (ref. 16). Germline transformants were made by injecting the clone of interest and the transposase plasmid, p π 25.7 wc (ref. 17), into γ ⁵⁰⁶ embryos¹⁸⁻²⁰. Transformants were crossed into the *Adh*¹⁶, *cn*; γ ⁵⁰⁶ background and established as homozygous stocks²¹.

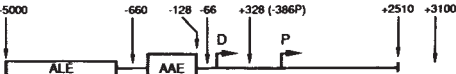
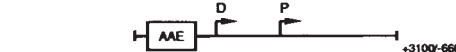
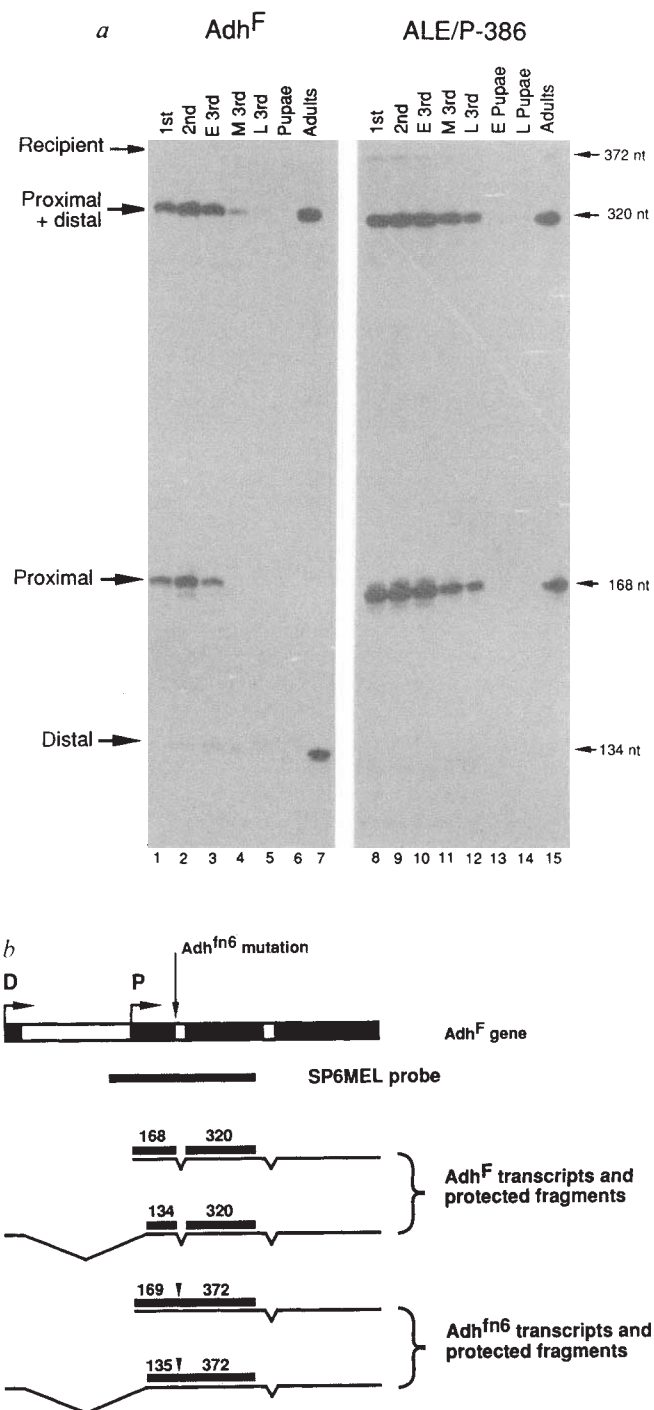
		Activity in adults		# of lines
		distal	proximal	
D-5000		+	-	5
ALE/P-386		N/A	3-30X↑	6
D-660/-128		10-20X↓	-	3
D-127/-63		3-5X↓	-	12
D-66/+53		N/A	3-8X↑	9
D+53/+328		+	-	9
D-660		+	-	5
5'3'AAE		+	-	6
5'P3'DP		+	10-15X↑	6
P-386		N/A	-	3

Fig. 3 The proximal promoter is aberrantly active in late third-instar larvae and in adults when the AAE and distal promoter are deleted. *a*, *Adh* transcript levels in *ALE/P-386* transformants were compared to wild-type controls at the indicated developmental stages by RNase mapping²². Relevant bands are marked at the side. *b*, Expected products from RNase mapping experiments using the *Adh*^F probe SP6MEL (ref. 21). The transcribed portion of the *Adh*^F gene is shown. Black boxes indicate exons; white boxes, introns. The extent of the complementary single-stranded RNA probe, SP6MEL, is indicated by the shaded bar below. Bracketed drawings show RNase-resistant hybrids which form between the probe and transcripts of the transformed (*Adh*^F) and endogenous (*Adh*^{fn6}) *Adh* genes. *Adh*^F transcripts give rise to three protected fragments: a 320-nucleotide (nt) fragment, common to proximal and distal transcripts; a 168 nt fragment, specific to proximal transcripts; and a 134 nt fragment, specific to correctly spliced distal transcripts. *Adh*^{fn6} transcripts are not spliced properly due to a mutation near the 3' splice site of the first small intron (marked by arrow)²³. Distal *Adh*^{fn6} transcripts yield fragments of 135 and 372 nt; proximal transcripts, fragments of 169 and 372 nt.

Methods. Total nucleic acid was isolated from staged organisms as previously described²¹ and hybridized to ³²P-labelled RNA probes complementary to *Adh* and α -tubulin transcripts. First, second, early third, mid third and late third-instar larvae were collected 30–42 h, 54–66 h, 66–78 h, 78–90 h and 90–102 h after egg laying, respectively. Early and late pupae were collected 24 and 72 h after the white pre-pupae stage. Adult males were collected 4 days after eclosion. RNA probes were made as described²⁴, except that bovine serum albumin was added to 10 μ g ml⁻¹. RNase mapping experiments were as described²², except hybridizations were at 37 °C, and the hybrids were digested for 30 min at 25 °C. The products were fractionated on 5% denaturing polyacrylamide gels and visualized by autoradiography.



of the position of the enhancers. The critical factor seems to be the position and activity of the distal promoter. The simplest interpretation of our data is that the distal promoter represses the proximal promoter by transcriptional interference.

Transcriptional interference was first shown to modulate promoter activity in *Escherichia coli*^{3,4,8}, and subsequently in vertebrates⁵⁻⁷. For example, induction of λ -prophage integrated upstream of the *E. coli gal* operon reduces transcription from the P_{g1} promoter⁴. Similarly, when the human α_1 globin gene is duplicated, transcription from the upstream gene inhibits transcription from the downstream gene⁶. Insertion of a transcription terminator between the two genes prevents this inhibitory effect⁶. Although the exact mechanism of transcriptional interference is not known, *in vitro* studies suggest that when

RNA polymerase passes through a downstream promoter it can destabilize interactions between that promoter and specific transcription factors⁷.

Here, we present evidence that transcriptional interference is important in the promoter switch of the *D. melanogaster Adh* gene. The proximal promoter is repressed in late third-instar larvae and adults apparently by readthrough transcription from the distal promoter. The proximal promoter seems to be sensitive to even low levels of transcription from the distal promoter (Fig. 4a, lanes 6–8), which may explain the rapidity of the switch from the proximal to distal promoter in third-instar larvae. Thus, the *Adh* promoter switch seems to be controlled largely or entirely by the stage-specific activation of the distal promoter which, in turn, may be regulated by the stage-specific activity

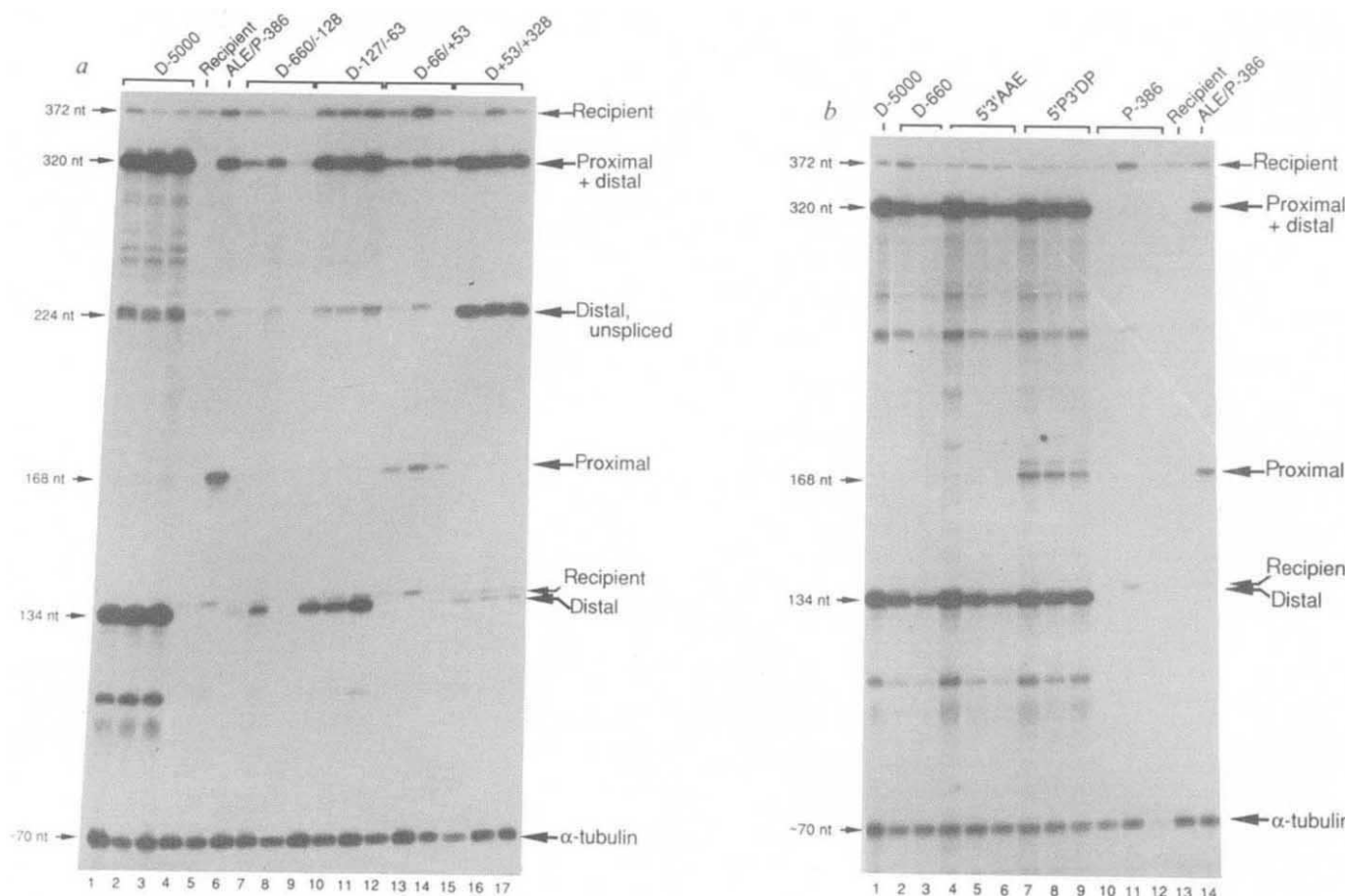


Fig. 4 The distal promoter is required to repress transcription from the proximal promoter in adults. Proximal and distal *Adh* transcripts from adult transformants were quantitated by RNase mapping. Each lane represents an independent transformed line of the indicated genotype. Relevant bands are marked at the side of the autoradiograms. *a*, In the absence of the AAE, the activity of the distal promoter was reduced 8- to 20-fold, but the proximal promoter was active at normal low levels (*D*-660/-128, lanes 6-8). Similarly, the activity of the distal promoter was reduced ~5-fold by the deletion of sequences just upstream of the distal TATA box, but the proximal promoter was unaffected (*D*-127/-63, lanes 9-11). The deletion of the sequence just downstream of the distal promoter did not affect the activity of either promoter (*D*+53/+328, lanes 15-17), although distal transcripts were not properly spliced due to the deletion of the 5' splice site of the large intron²³. In contrast, when the distal promoter's TATA box and transcription initiation site were deleted, the proximal promoter was expressed at levels 2- to 10-times higher than normal (*D*-66/+53, lanes 12-14). *b*, Proximal transcripts from 5'P3'DP were detected at levels 10- to 15-fold higher than controls (compare lanes 7-9 to lanes 1-3). Note that both the upstream and downstream proximal promoters contribute to the pool of proximal transcripts. The downstream proximal promoter, however, is active at extremely low levels in controls (lanes 1-3). Therefore the vast majority of the proximal transcripts must initiate at the upstream proximal promoter.

Methods. Total nucleic acids from adult males were treated as described in Fig. 3 legend.

of the transcription factors that bind to it and the AAE (ref. 9 and V.C., D. Falb and T.M., unpublished observations).

Only one other example of a stage-specific promoter switch has been studied in detail. In the chicken β -globin locus the ϵ -globin gene is expressed in embryonic erythroid cells, and the closely linked β -globin gene is expressed in adult erythroid cells^{10,11}. Recent studies have shown that the ϵ -globin gene is inappropriately expressed in adult cells if the β -globin gene promoter is deleted¹². The promoter switch in both the *Adh* and globin systems seem to be regulated by the stage-specific activation of an adult promoter. In the globin locus, however, the ϵ -globin gene is repressed in adult cells by competing with the β -globin promoter for interactions with a single erythroid cell-specific enhancer¹². In contrast, the distal *Adh* promoter represses the proximal promoter by transcriptional interference. Thus two distinct mechanisms appear to regulate promoter switching in these two systems.

We thank Bob Cohen, Anne Ephrussi, Sean Munro, Mark Ptashne and members of the Maniatis lab for helpful comments on the manuscript. The work was supported by a grant to T.M. from the NIH.

Received 28 October; accepted 30 November 1988.

- Benayati, C., Spoerel, N., Haymerle, H. & Ashburner, M. *Cell* **33**, 125-133 (1983).
- Savakis, C., Ashburner, M. & Willis, J. H. *Dev Biol.* **114**, 194-207 (1986).
- Hausler, B. & Somerville, R. L. *J. molec. Biol.* **127**, 353-356 (1979).
- Adhya, S. & Gottesman, M. *Cell* **29**, 939-944 (1982).
- Cullen, B. R., Lomedico, P. & Ju, G. *Nature* **307**, 241-245 (1984).
- Proudfoot, N. J. *Nature* **322**, 562-565 (1986).
- Bateman, E. & Paule, M. R. *Cell* **54**, 985-992 (1988).
- Gay, N. J., Tybulewicz, V. L. & Walker, J. E. *Biochem. J.* **234**, 111-117 (1986).
- Heberlein, U. & Tjian, R. *Nature* **331**, 410-415 (1988).
- Bruns, G. A. & Ingram, V. M. *Phil. Trans. R. Soc. B* **266**, 225-305 (1973).
- Brown, J. L. & Ingram, V. M. *J. biol. Chem.* **249**, 3960-3972 (1974).
- Choi, O. B. & Engle, J. D. *Cell* **55**, 17-26 (1988).
- Posakony, J. W., Fischer, J. A. & Maniatis, T. *Cold Spring Harb. Symp. quant. Biol.* **50**, 515-520 (1985).
- Spradling, A. C. & Rubin, G. M. *Cell* **34**, 47-57 (1983).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratories, New York, 1982).
- Rubin, G. M. & Spradling, A. C. *Nucleic Acids Res.* **11**, 6341-6351 (1983).
- Karess, R. E. & Rubin, G. M. *Cell* **38**, 135-146 (1984).
- Rubin, G. M. & Spradling, A. C. *Science* **218**, 348-353 (1982).
- Spradling, A. C. & Rubin, G. M. *Science* **218**, 341-347 (1982).
- Goldberg, D. A., Posakony, J. W. & Maniatis, T. *Cell* **34**, 58-73 (1983).
- Fischer, J. A. & Maniatis, T. *EMBO J.* **5**, 1275-1289 (1986).
- Zinn, K., DiMaio, D. & Maniatis, T. *Cell* **34**, 865-879 (1983).
- Benayati, C., Place, A., Wang, N., Pentz, E. & Sofer, W. *Nucleic Acids Res.* **10**, 7261-7272 (1982).
- Melton, D. et al. *Nucleic Acids Res.* **12**, 7035-7056 (1984).

Mutation rates differ among regions of the mammalian genome

Kenneth H. Wolfe, Paul M. Sharp* & Wen-Hsiung Li†

Department of Genetics, Trinity College, Dublin 2, Ireland

† Center for Demographic and Population Genetics, University of Texas, PO Box 20334, Houston, Texas 77225, USA

In the traditional view of molecular evolution, the rate of point mutation is uniform over the genome of an organism and variation in the rate of nucleotide substitution among DNA regions reflects differential selective constraints^{1,2}. Here we provide evidence for significant variation in mutation rate among regions in the mammalian genome. We show first that substitutions at silent (degenerate) sites in protein-coding genes in mammals seem to be effectively neutral (or nearly so) as they do not occur significantly less frequently than substitutions in pseudogenes. We then show that the rate of silent substitution varies among genes and is correlated with the base composition of genes and their flanking DNA. This implies that the variation in both silent substitution rate and base composition³ can be attributed to systematic differences in the rate and pattern of mutation over regions of the genome. We propose that the differences arise because mutation patterns vary with the timing of replication of different chromosomal regions in the germline. This hypothesis can account for both the origin of isochores in mammalian genomes⁴ and the observation⁵ that silent nucleotide substitutions in different mammalian genes do not have the same molecular clock.

Early estimates of the substitution rates in pseudogenes and at silent sites in protein-coding sequences suggested that pseudogenes evolve more rapidly which implies that silent sites are under some selective constraint^{6,7}. However, these studies used very limited data and rather simple methods of estimating rates. Here we use two approaches to compare the rate of substitution in functionless DNA (that is, in pseudogenes and intergenic regions) with that at silent sites in genes; the latter is measured in terms of the number of nucleotide substitutions per fourfold degenerate site (designated K_4). First, we compare the divergences in pseudogenes and in functional genes across the same species pair. If silent sites in functional genes are selectively constrained, then they should have accumulated fewer substitutions.

By far the most substantial data set available is for DNA sequences from humans and Old World monkeys (Table 1). The mean degree of divergence at silent sites is clearly not lower than that in pseudogenes. There are rate differences among genes, but for only one gene (β -globin) out of 13 is the silent rate significantly lower than the pseudogene rate, suggesting that the β -globin rate is an extreme random deviate. Nevertheless, perhaps a better method is to compare a recently generated pseudogene with its functional homologue from the same species, using the functional gene from a second species as a reference. This method can be applied to both processed and unprocessed pseudogenes, but is more accurate in the former case because processed pseudogenes can be assumed to become functionless as soon as they are formed. The three LDH-A processed pseudogenes in mouse⁸ provide excellent data for this approach because they are reasonably large and because they were produced after the mouse-rat split, so the rat LDH-A gene can be used as a reference. The ratios of the substitution rates in the three LDH-A pseudogenes to the rate at silent sites in the functional gene are 1.1, 0.8 and 0.7. Again, there is little, if any, evidence for a significantly higher rate of substitution in pseudogenes than at silent sites. We therefore conclude that most silent (fourfold degenerate) sites are effectively neutral, and that their substitution rate is a good measure of the mutation

Table 1 Rates of nucleotide substitution at silent sites and in functionless DNA for human versus Old World monkey sequences

Gene	$K_4 \pm \text{s.e.m.}$	GC [*]	L_4 [†]
α_1 -Antitrypsin	0.077 ± 0.022	73.6	184
Apolipoprotein A-I	0.061 ± 0.025	88.4	126
Apolipoprotein E	0.062 ± 0.021	89.1	175
Chorionic gonadotrophin- β	0.065 ± 0.027	81.3	99
Erythropoietin	0.110 ± 0.035	71.7	110
β -Globin	0.039 ± 0.028	68.2	54
Insulin	0.130 ± 0.054	82.3	62
Metallothionein I	0.125 ± 0.113	85.5	21
Metallothionein II	0.096 ± 0.072	87.0	23
β -Myosin heavy chain	0.179 ± 0.056	86.5	97
Pepsinogen A	0.108 ± 0.026	73.7	194
Transforming growth factor- β	0.064 ± 0.019	83.5	203
Triose phosphate isomerase	0.112 ± 0.033	59.8	128
Total for silent sites	0.089 ± 0.009	78.7	1,474
η -Globin pseudogene locus	0.074 ± 0.006	42.9	2,071
η - δ Globin intergenic DNA	0.076 ± 0.006	38.7	2,771
Total for functionless DNA	0.075 ± 0.004	40.5	4,842

K_4 (the corrected number of substitutions per fourfold degenerate site) was calculated as described in Fig. 1. References to the original DNA sequence data are available on request.

* G + C content at fourfold degenerate sites (%).

† Number of fourfold degenerate sites compared.

rate, because under selective neutrality the substitution rate is expected to be equal to the mutation rate¹.

Next we focus on the variation among genes in the substitution rate at silent sites (K_4) and show that it is correlated with their G + C content (GC₄). We compare DNA sequences between mouse and rat, first concentrating on the larger genes (those with over 175 silent sites) (Fig. 1a). For the 17 genes with GC₄ $\geq$ 50%, K_4 clearly decreases as GC₄ increases (correlation coefficient $r = -0.73$, $P < 10^{-4}$); the highest K_4 is nearly twice the lowest value. On the other hand, there is a significant positive correlation between substitution rate and GC₄ in the A + T-rich genes, though the number of points is small ($r = +0.82$; $P < 0.05$). This latter observation suggests that the low substitution rates in genes with a high G + C content cannot be simply explained by assuming that G and C nucleotides mutate less frequently than A and T nucleotides. On reflection, it is difficult to think of any other reasonable relationship between K_4 and GC₄, although this has not previously been reported. In fact, others have suggested either no systematic variation (in silent substitution rate) among genes⁹, or a simple linear relationship between substitution rate and G + C content^{10,11}. When smaller genes are included (making a total of 88 genes), the same peaked relationship as that shown in Fig. 1a is seen, but the correlation coefficients are decreased because of the stochastic effects (particularly on K_4) of short genes. The above comparison requires no knowledge of the date of speciation, the data set used is large, and the K_4 values are fairly large but far from saturation. The mouse-rat comparison is preferable to a human-rodent comparison because of the shorter timescale involved; some human genes differ considerably in base composition¹² or even in chromosomal map position¹³ from their rodent homologues.

The variance of GC₄ in the 23 large rodent genes (Fig. 1a) is more than ten times that expected by chance (under a binomial distribution with mean 58.6%), suggesting that the factors determining G + C content are not uniform for all genes. Of course, one might argue that the GC₄ value in any gene is at a particular optimum and that the observed relationship between K_4 and GC₄ is a consequence of selection maintaining that GC₄ value in the face of a mutation rate that is uniform across the genome. But this would require strong selective constraints, whereas we have shown here that silent sites seem to be effectively neutral. Furthermore, the direction of selection on G + C-rich genes

* To whom correspondence should be addressed.

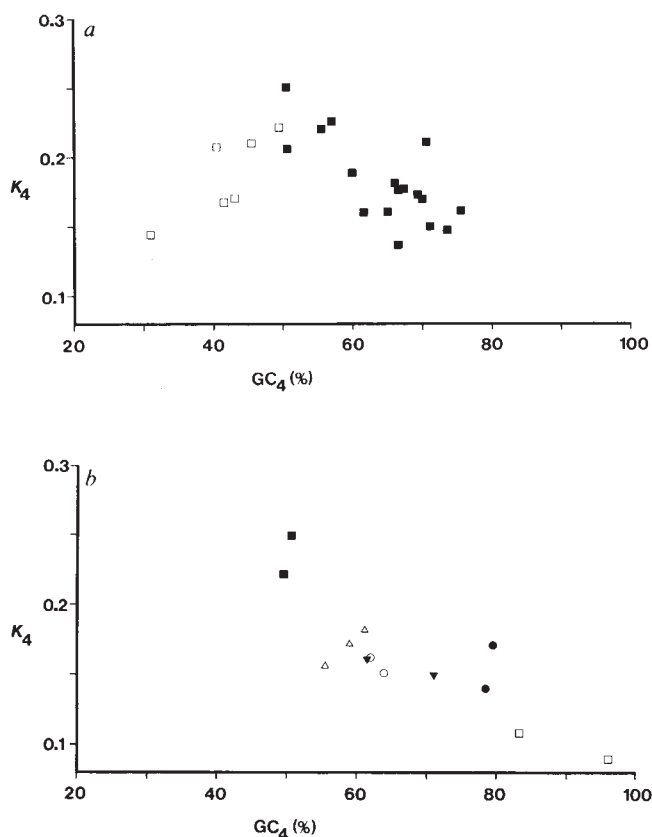


Fig. 1 *a*, Plot of silent substitution rate (K_4) versus G+C content at silent sites (GC_4) for 23 genes compared between mouse and rat. Only large genes (with ≥ 175 fourfold degenerate sites) are compared. ■, 17 genes for which $GC_4 \geq 50\%$ (encoding alcohol dehydrogenase-1, α -fetoprotein, CD4 antigen, c-mos and c-myc oncogenes, creatine kinase M, cytochromes P_{450} and P_{3450} , β -glucuronidase, malic enzyme, mitochondrial aspartate aminotransferase and malate dehydrogenase, neural cell adhesion molecule NCAM-140, neurofilament M, oestrogen receptor, protein disulphide isomerase and renin); □, 6 genes with $GC_4 < 50\%$ (encoding albumin, α_2 -amylase, glucocorticoid receptor, leukocyte common antigen (Ly-5, T200), ornithine decarboxylase and UDP-glucuronosyltransferase). Full references for the sequences are available from the authors on request; all are taken from GenBank (release 50) or the recent literature. K_4 was calculated by the method of Tajima and Nei³⁰, which requires no assumptions about the base composition of the sequences compared, other than that they be at equilibrium. Any codons at which mouse and rat differed by more than one nucleotide substitution were ignored. This avoided (1) the need to apply weights to the different mutational pathways possible between multiply-substituted codons, and (2) any regions within a gene in which the protein sequences of the two species are considerably different and might not share a common origin (for example as a result of insertions, deletions or alternative splicing). If, as we assume, the sequences being compared have diverged only through a process of random point mutation, ignoring these codons should not alter the synonymous substitution rate calculated. *b*, Relationship between K_4 and GC_4 for several groups of physically linked genes: □, metallothioneins I and II (5 kb apart); ■, albumin and α -fetoprotein (14 kb); ▼, cytochromes P_{450} and P_{3450} (<25 kb); ●, γ_A and γ_D crystallins (35 kb); ○, Ly-2 and Ly-3 antigens (36 kb); △, immunoglobulin heavy chains C_δ , C_{γ_3} and C_ϵ (50 kb and 90 kb).

would have to be opposite to that on A+T-rich genes. Thus our finding that the substitution rate and the base composition of silent sites vary together in a systematic way is most simply explained by supposing that the pattern of mutation is different for different genes. Most germline mutations are thought to arise from misincorporation errors made by the DNA replication apparatus^{14,15}. It has been demonstrated that different genes replicate at different stages of the cell cycle in differentiated cells¹⁶, and this is presumably also true in the germline. The number and type of replication errors are likely to vary during the cell cycle if the chemical environment in the nucleus changes. In fact, the abundances (both relative and absolute) of free dNTPs in the nucleus change with time¹⁷, as do the activities of the DNA polymerase enzymes and their accessory proteins¹⁸. We have examined a theoretical model (to be detailed elsewhere) of the relationship between the mutation rate and the G+C content of both the nucleotide precursor (dNTP) pools and the template DNA. This model is based on a simple kinetic description of DNA replication¹⁹, as applied to the replication of long double-stranded DNA molecules. It can explain the occurrence of the maximum mutation rate at an intermediate value of GC_4 , provided that GC_4 is positively correlated with the proportion of dGTP plus dCTP in the dNTP pools. This assumption seems plausible as we are considering only effectively neutral sites in DNA. There is also some experimental evidence to support this: as S-phase progresses, the G+C content of both the dNTP pools and the replicating DNA decreases^{16,17,20}. As K_4 is a consequence of the mutation rate, under this model the relationship seen in Fig. 1a can be generated in the absence of selection.

Replication origins in mammalian DNA are located 50–300 kilobases (kb) apart, and clusters of 12–100 adjacent origins tend to replicate simultaneously¹⁶. Thus, regions of DNA of 1,000 kb or more could form temporal units of replication, and all DNA within a unit would be expected to show the same pattern and rate of mutation. Thus, under our model, genes which are physically close to each other in the genome should

have similar values of K_4 and GC_4 . This is indeed the case for those genes for which data are available (Fig. 1b): several pairs of genes which are the results of ancient tandem duplications show paired results for K_4 and GC_4 , as do genes of the IgH constant-region locus. These IgH genes span 140 kb on mouse chromosome 12, and are known to be part of a single replicon²¹.

These replication units seem to correspond to the 'isochores' (large blocks of DNA with homogeneous base compositions) found experimentally in mammalian nuclear DNA⁴, and probably also to high-resolution Giemsa chromosomal bands¹⁶. We therefore propose that isochores arise as a result of the synchronous replication of megabase stretches of DNA under varying dNTP pool conditions. Although our model based on variation in the dNTP precursor pools can provide a simple explanation for the observed variation of mutation rates and patterns around the genome, it is of course not the only possible explanation. For example, Filipinski²² has proposed that isochores are formed as a consequence of the repair of DNA in different types of chromatin by different DNA polymerase enzymes, and there is now some experimental evidence for between-gene differences in efficiency of DNA repair²³. Our observations could also be explained if DNA is replicated by several distinct DNA polymerase holoenzymes with different error propensities. But recent data indicate that there is only one replicative polymerase complex for mammalian nuclear DNA^{18,24}.

We conclude that much of the intragenomic variation in silent substitution rate and base composition in mammals results from variation in the process of mutation, rather than from natural selection^{25,26}. The implication that silent sites in mammalian genes are effectively neutral contrasts with prokaryotic genes, in which differences in translational efficiency lead to selective differences between codons²⁷ and thus to lower silent substitution rates in genes that are highly expressed². Our assertion is supported by several arguments, however. First, the base composition of silent sites in mammalian genes is strongly correlated with that of introns and flanking sequences of the same gene³,

suggesting that it is determined largely by local mutation pressures. In addition, it has recently been suggested on the basis of analysis of dinucleotide frequencies that codon usage in human genes is determined by context-dependent mutational patterns²⁸. Second, any suggestion of tissue-specific or expression level-related patterns in mammalian codon usage must be set against the observation that the human α - and β -globin genes have very different codon usage. Third, as the effective population sizes of mammalian species are quite small²⁹, it would require large selective differences for selection to be effective, but it is unlikely that mammals have sufficient reproductive excess to allow such selection at a very large number of codons. In summary, we propose that, in the mammalian genome, variation in the silent substitution rate and variation in base composition are two facets of the same phenomenon.

We thank D. W. Nebert for cytochrome P_{450} mapping data. This study was supported in part by the European Community BAP (to P.M.S.) and the NIH (to W.H.L.).

Received 5 July; accepted 7 December 1988.

- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).
- Sharp, P. M. & Li, W.-H. *Molec. biol. Evol.* **4**, 222-230 (1987).
- Aota, S. & Ikemura, T. *Nucleic Acids Res.* **14**, 6345-6355 (1986).
- Bernardi, G. *et al. Science* **228**, 953-958 (1985).
- Li, W.-H., Tanimura, M. & Sharp, P. M. *J. molec. Evol.* **25**, 330-342 (1987).
- Li, W.-H., Gojobori, T. & Nei, M. *Nature* **292**, 237-239 (1981).
- Miyata, T. & Hayashida, H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5739-5743 (1981).
- Fukasawa, K. M. *et al. Genetics* **115**, 177-184 (1987).
- Miyata, T. *et al. J. molec. Evol.* **19**, 28-35 (1982).
- Filipski, J. *J. theor. Biol.* **134**, 159-164 (1988).
- Smithies, O., Engels, W. R., Devereux, J. R., Slightom, J. L. & Shen, S.-h. *Cell* **26**, 345-353 (1981).
- Mouchiroud, D. & Gautier, C. *Molec. biol. Evol.* **5**, 192-194 (1988).
- Nadeau, J. H. & Taylor, B. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 814-818 (1984).
- Friedberg, E. C. *DNA Repair* (Freeman, New York, 1985).
- Topal, M. D. & Fresco, J. R. *Nature* **263**, 285-289 (1976).
- Holmquist, G. P. *Am. J. hum. Genet.* **40**, 151-173 (1987).
- Leeds, J. M., Slabaugh, M. B. & Mathews, C. K. *Molec. cell Biol.* **5**, 3443-3450 (1985).
- Kelly, T. & Stillman, B. (eds) *Cancer Cells 6: Eukaryotic DNA Replication* (Cold Spring Harbor Laboratory, New York, 1988).
- Fersht, A. R. & Knill-Jones, J. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4251-4255 (1981).
- Holmquist, G., Gray, M., Porter, T. & Jordan, J. *Cell* **31**, 121-129 (1982).
- Brown, E. H. *et al. Molec. cell Biol.* **7**, 450-457 (1987).
- Filipski, J. *FEBS Lett.* **217**, 184-186 (1987).
- Bohr, V. A., Phillips, D. H. & Hanawalt, P. C. *Cancer Res.* **47**, 6426-6436 (1987).
- Prelich, G. & Stillman, B. *Cell* **53**, 117-126 (1988).
- Bernardi, G. & Bernardi, G. *J. molec. Evol.* **24**, 1-11 (1986).
- Gillespie, J. H. *Genetics* **113**, 1077-1091 (1986).
- Ikemura, T. *Molec. biol. Evol.* **2**, 13-34 (1985).
- Hanai, R. & Wada, A. *J. molec. Evol.* **27**, 321-325 (1988).
- Nei, M. & Graur, D. *Evol. Biol.* **17**, 73-118 (1984).
- Tajima, F. & Nei, M. *Molec. biol. Evol.* **1**, 269-285 (1984).

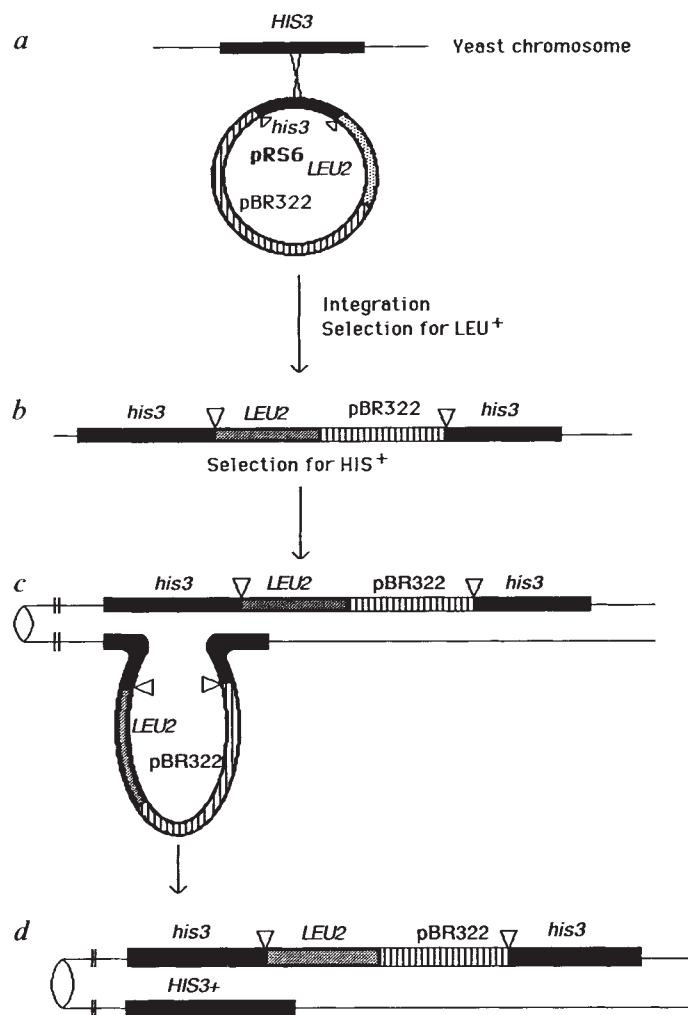


Fig. 1 Plasmid pRS6 which contains an internal fragment of the *HIS3* gene was cut within the internal *his3* fragment and integrated into the genome at the *HIS3* locus **a**. This creates a duplication of the *his3* gene in which one allele is deleted for its 3' end and the other for its 5' end **b**. The two alleles share about 400 base pairs of homology and thus can recombine with each other to revert to the *HIS3*⁺ allele. As shown previously *HIS3*⁺ recombinants do not arise by plasmid excision or by unequal sister chromatid exchange¹³. In these studies the frequency of plasmid excision was determined by putting a yeast origin of replication sequence onto the integrating plasmid. Excised plasmids could be recovered but were 100-fold less frequent than the frequency of *HIS*⁺ formation which suggests that plasmid excision does not occur in the majority of recombinants. Reciprocal products expected to result from sister chromatid exchange were analysed by Southern blotting. The pattern characteristic for sister chromatid exchange was not found in any of 25 events examined. Thus, as a likely alternative mechanism conversion between sister chromatids **c** is suggested which result in deletion of the integrated plasmid on one chromatid **d**. After segregation of the two chromatids the *HIS*⁺ recombinant should show a *HIS*⁺ *leu*⁻ phenotype which is found in 99% of all *HIS*⁺ recombinants¹³. Thus, it is proposed, as a possible model, that a double strand break initiates the recombination event which is extended by an exonuclease to a gap. This gap is repaired by gene conversion from the sister chromatid as donor¹³.

Nonmutagenic carcinogens induce intrachromosomal recombination in yeast

Robert H. Schiestl

Department of Biology, University of Rochester, Rochester, New York 14627, USA
Prairie Biological Research Ltd, 10515-36A Avenue, Edmonton, Alberta T6J 2H7, Canada
GeneBioMed, Inc., PO Box 18121, 12 Corners Station, Rochester, New York 14618, USA

To identify environmental carcinogens there is a need for inexpensive and reliable short-term tests¹, but certain human or animal carcinogens are persistently undetectable as mutagens with the Ames assay²⁻⁵ or with other short-term tests currently in use^{6,7}. Thus there is a need for short-term tests which detect carcinogens missed by the Ames assay¹. Because of the association of carcinogenesis with genome rearrangement⁸⁻¹², a system screening for intrachromosomal recombination resulting in genome rearrange-

ment has been constructed in *Saccharomyces cerevisiae*¹³. Evaluation of this system shows inducibility by a variety of carcinogens³⁻⁷ not detectable by the Ames assay or various other short-term tests. In the light of these results it is tempting to speculate that 'nongenotoxic carcinogens' are in fact genotoxic but, in the past, the tools to measure the genetic alterations they induce have been inappropriate.

Table 1 Carcinogens which are detectable with the Ames assay

Carcinogen	Response in DEL (concentration range in $\mu\text{g ml}^{-1}$)	Response in ICR (concentration range in $\mu\text{g ml}^{-1}$)	Concentration in $\mu\text{g ml}^{-1}$	HIS ⁺ per 10 ⁴ cells (colonies counted)	ADE ⁺ per 10 ⁵ cells (colonies counted)
MMS ¹⁸	Positive (1.3–260)	Positive (1.3–260)	0 260	0.97 (87) 129 (137)	0.45 (20) 62 (654)
EMS ¹⁹	Positive (234–2,340)	Positive (234–2,340)	0 2,340	0.97 (87) 25.3 (1,428)	0.45 (20) 28.9 (155)
NQO ²⁰	Positive (0.1–2)	Positive (0.1–2)	0 2	0.97 (87) 55.6 (562)	0.45 (20) 106 (252)
Nitrogen mustard ³	Positive (0.1–100)	Positive (0.1–100)	0 100	0.79 (73) 176 (84)	1.3 (120) 460 (22)
Epichlorohydrine ¹⁸	positive (100–400)	Positive (100–400)	0 400	0.68 (76) 48 (2,976)	0.58 (65) 50 (620)
Aflatoxin B ₁ ¹⁸	Positive (2.5–50)	Positive (2.5–50)	0 50	0.68 (76) 12.5 (277)	0.58 (65) 3.15 (70)
1–2 Dibromoethane ²²	Positive (22–4,400)	Positive (22–4,400)	0 4,400	1.2 (127) 3.8 (31)	3.3 (328) 12.2 (10)
Dimethylhydrazine ^{18,23}	Positive (10–5,000)	Positive (10–5,000)	0 5,000	0.78 (75) 3.9 (131)	2.5 (245) 4.7 (16)

The influence of MMS exposure (vol/vol) on the frequency of deletions (DEL) and interchromosomal recombination (ICR) was determined. The diploid strain RS112 (MATa/ α *ura3-52/ura3-52 leu2-3,112/leu2-Δ98 trp5-27/TRP5 arg4-3/ARG4 ade2-40/ade2-101 ilo1-92/ILV1 HIS3::pRS6/his3-Δ200 LYS2/lys2-801*) was used and contains the DEL system on one homologue (*HIS3::pRS6*) and a deletion of the entire open reading frame of *HIS3* on the other homolog (*his3-Δ200*)²⁷. Single colonies were picked from YPAD medium²⁸ and were inoculated into 5–25 ml of synthetic complete medium lacking leucine (SC-LEU) and grown for 24 hours at 30 °C under constant shaking. Cells were counted and the cell density was adjusted to 2×10^6 cells ml⁻¹ in fresh SC-LEU medium. The medium containing the cells was distributed in aliquots of 5 ml each into disposable 15-ml tubes. The agent to be tested was added, the tubes sealed, and the cells were incubated for 17 hours at 30 °C under constant shaking. Cells were pelleted in a clinical table-top centrifuge. After washing the cells twice with sterile distilled water or a 10% solution of sodium thiosulphate to inactivate the agent, cells were resuspended in 0.5–1 ml of sterile distilled water, transferred to a glass tube and sonicated to disperse any clumps. Cells were counted and appropriate numbers were plated onto SC medium to determine the number of survivors, onto SC-HIS medium to score for DEL events and onto SC-ADE medium to determine the number of ICR events. The cells did not contain any clumps or tetrads. Colonies were counted after two to three days of incubation at 30 °C. Numbers of recombinants were calculated per numbers of survivors. Shown is the concentration of the agent, the number of generations obtained after 17 hours of incubation, the survival and DEL as well as ICR frequencies. Data derived from less than five colonies were not included. The same control was used for MMS, EMS and NQO as well as for epichlorohydrin and aflatoxin exposure respectively. A minimum increase of twofold over the spontaneous frequency in a dose dependent manner has been regarded as evidence for inducibility. Each experiment was repeated at least three times and the results obtained were highly reproducible, two plates were used for each system and each concentration. The same qualitative result was obtained with other diploid as well as haploid strains containing the DEL system. MMS and EMS were purchased from VWR Scientific and purified by vacuum distillation at 10–15 mm Hg pressure before use. NQO, nitrogen mustard, dibromoethane (ethylene dibromide), epichlorohydrine, aflatoxin and dimethylhydrazine were obtained from Sigma.

There is a large body of data establishing that animal and human carcinogens without apparent genotoxic activity exist⁶. This is the basis for the concern that there may be a considerable number of nonmutagenic carcinogens in our environment that are not detected by the conventional short term tests. As models for carcinogenesis of 'nongenotoxic carcinogens' are largely missing, a significant regulatory dilemma in their assessment of risk to humans exists¹⁴.

There is evidence that substantial genome rearrangements are associated with cancer^{8–12}. Furthermore deletions in recessive oncogenes have been implicated in carcinogenesis such as retinoblastoma^{15,16}. Because of the association of genome rearrangement with cancer, a system selecting for intrachromosomal recombination which results in genome rearrangement has been constructed in the yeast *S. cerevisiae*¹³ (Fig. 1). A plasmid with an internal fragment of the *HIS3* gene has been integrated at the *HIS3* locus yielding an integrative disruption of the *HIS3* gene. This resulted in two copies of the *HIS3* gene each having one terminal deletion. The most likely mechanism for reversion seems to be sister chromatid conversion between the two *his3*⁻ alleles¹³. Because this recombination event deletes the integrated plasmid (Fig. 1) it is termed a deletion (DEL) event.

The DEL system is under different genetic control than the previously described recombination systems interchromosomal recombination (ICR) and meiotic recombination. It has been shown that mutations in the DNA repair gene *RAD1* affect the DEL recombination but have no effect on ICR or meiotic recombination¹⁷. This suggests that the mechanism for DEL recombination differs from that for ICR as well as meiotic recombination.

To determine whether DEL is inducible by carcinogens the diploid strain RS112 was used in which all recombination events have to occur between the two *his3*⁻ alleles on one chromosome because the entire *HIS3* gene is deleted on the homologous chromosome. RS112 is also heteroallelic for *ade2* to measure interchromosomal recombination (ICR) between homologous chromosomes and the two recombination systems can be compared in their performance to detect carcinogens.

As mutagenic carcinogens methyl methanesulphonate (MMS)^{3,18}, ethyl methanesulphonate (EMS)^{3,19}, 4-nitroquinoline-*N*-oxide (NQO)^{3,7,20}, nitrogen mustard³, epichlorohydrine^{5,7,18} and aflatoxin B₁^{3,18} were used (see Table 1 for procedure) and gave strong inductions with the DEL system (HIS⁺; 20–220-fold, Table 1) as well as with the ICR system (ADE⁺, also see ref. 21). Ethylene dibromide²² is weakly positive with the Ames assay³, and dimethylhydrazine^{18,23} which has been reported negative³ or weakly positive⁵ with the Ames assay, both induce DEL recombination but ICR to a lesser extent (Table 1).

The carcinogens formaldehyde, safrole, ethionine, urethane, auramine, methylene chloride, carbon tetrachloride, cadmium chloride and sulfate, aniline, aminotriazole, acetamide, thioacetamide, thiourea, dichlorodiphenyldichloroethylene (DDE) and ethylenethiourea are negative with all five strains currently used in the Ames assay^{3,5,22}, yet they have been shown to be animal or suspected human carcinogens (see Table 2 for references). All of these carcinogens induce the DEL system (2–13-fold; see Table 2).

In an international collaborative study with coded chemicals evaluating 30 different short term assays in more than 50 different laboratories⁷ the carcinogens auramine, safrole, urethane,

Table 2 Agents which are not detectable as 'carcinogens' in the Ames assay

Agent*	Response in DEL (concentration range in $\mu\text{g ml}^{-1}$)	Response in ICR (concentration range in $\mu\text{g ml}^{-1}$)	Concentrations ($\mu\text{g ml}^{-1}$)	HIS ⁺ per 10 ⁴ cells (colonies counted)	ADE ⁺ per 10 ⁵ cells (colonies counted)
Formaldehyde ^{23,29}	Positive (15–70)	Positive (15–50)	0 50	2.3 (212) 29.5 (109)	0.49 (45) 5.4 (6)
Safrole ^{7,18,23}	Positive (2,000–4,000)	Negative (up to 2,000)	0 2,000	1.1 (129) 2.4 (117)	0.50 (55) 0.82 (40)
Ethionine ^{7,18,30}	Positive (1,000–20,000)	Negative (up to 20,000)	0 20,000	1.05 (108) 5.4 (250)	0.66 (68) 0.80 (47)
Urethane ^{7,18,23}	Positive (20,000–40,000)	Negative (up to 20,000)	0 20,000	0.97 (99) 4.8 (478)	0.45 (47) 0.40 (20)
Auramine ^{18,31}	Positive (1,200–1,500)	Negative (up to 1,200)	0 1,200	1.2 (127) 2.64 (78)	3.3 (328) 1.7 (5)
Methylene chloride ^{22,29,31}	Positive (10,000–12,000)	Negative (up to 12,000)	0 12,000	0.95 (119) 3.6 (109)	1.1 (693) 0.98 (5)
Carbon tetrachloride ^{18,23,29}	Positive (4,000–8,000)	Negative (up to 4,000)	0 4,000	1.05 (108) 2.0 (99)	0.66 (68) 0.61 (6)
Cadmium chloride ^{18,31}	Positive (100–700)	Negative (up to 400)	0 400	0.94 (95) 3.5 (145)	0.49 (55) 0.30 (8)
Cadmium sulphate ^{18,31}	Positive (100–700)	Negative (up to 400)	0 200	0.94 (95) 2.6 (516)	0.49 (55) 0.2 (5)
Aniline ²²	Positive (4,000–7,000)	Positive (at 7,000)	0 7,000	1.18 (134) 7.9 (321)	1.34 (153) 4.9 (20)
3-Aminotriazole ¹⁸	Positive (40,000–60,000)	Negative (up to 60,000)	0 60,000	1.13 (91) 3.1 (798)	1.08 (86) 0.7 (19)
Acetamide ¹⁸	Positive (20,000–40,000)	Negative (up to 40,000)	0 40,000	1.13 (91) 5.04 (782)	1.08 (86) 0.45 (7)
Thioacetamide ¹⁸	Positive (10,000–20,000)	Negative (up to 20,000)	0 20,000	1.13 (91) 6.28 (1,432)	1.08 (86) 1.14 (26)
Thiourea ¹⁸	Positive (10,000–30,000)	Positive (at 30,000)	0 30,000	1.13 (91) 25.3 (76)	1.08 (86) 20 (6)
DDE ²²	Positive (100–5,000)	Positive (at 5,000)	0 5,000	1.99 (112) 14.1 (97)	0.57 (32) 4.3 (5)
Ethylenethiourea ^{7,18}	Positive (20,000–40,000)	Positive (at 40,000)	0 40,000	0.97 (85) 3.1 (576)	0.56 (49) 1.5 (32)
Methionine	Negative (up to 40,000)	Negative (up to 40,000)	0 40,000	0.95 (119) 0.64 (66)	1.1 (693) 0.69 (358)

The experiment was carried out as described in Table 1. The controls for the agents ethionine and carbon tetrachloride, for cadmium chloride and sulphate, for 3-aminotriazole and acetamide, and for thioacetamide and thiourea were the same. Formaldehyde, methylene chloride (dichloromethane) and carbon tetrachloride were obtained from J. T. Baker Chemical Company. Safrole, ethionine, auramine, cadmium chloride, cadmium sulfate, aniline, 3-aminotriazole, acetamide, thioacetamide, thiourea and methionine were from Sigma. Urethane, DDE (2,2-bis[4chlorophenyl]-1,1-dichloroethylene) and ethylenethiourea (2-Imidazolidinethione) were from Aldrich.

* References are given for the carcinogenicity of carcinogenic agents.

ethionine, ethylenethiourea, and 3-aminotriazole gave very poor inducibilities with all short term tests.⁷ In the collaborative study as many short term tests were positive with the amino acid methionine as were positive with the carcinogen ethionine and only very few tests could differentiate between the two compounds⁷. After incubation in up to 40 mg ml⁻¹ of methionine no increase of the frequency of DEL was detectable (Table 2), whereas ethionine gave already a twofold increase at a concentration of 1 mg ml⁻¹ (Table 2). Thus the DEL assay can readily differentiate between methionine and ethionine.

Formaldehyde, safrole, ethionine, urethane, 3-aminotriazole and ethylenethiourea have been shown to induce chromosome mal-segregation in fungi which may be caused by a nongenotoxic mechanism²⁴. Chromosome mal-segregation associated with DEL events would show a HIS⁺ LEU⁺ phenotype¹³. No evidence for an increase of the frequency of HIS⁺ LEU⁺ colonies was found. Therefore chromosome mal-segregation does not contribute significantly to the carcinogen induced frequency of HIS⁺ formation.

In comparing the performance of the two systems, DEL and ICR, it will be noticed that the mutagenic carcinogens induce both systems well (Table 1). The nonmutagenic carcinogens on the other hand induce the DEL system much better than the ICR system (Table 2), and the carcinogens safrole, ethionine, urethane, auramine, methylene chloride, carbon tetrachloride, cadmium sulfate and chloride, aminotriazole and acetamide would not have been detectable with the ICR system. The

deletion of the integrated plasmid (see Fig. 1), which is inducible by the 'nongenotoxic carcinogens' used, indicates a genotoxic activity of these carcinogens.

In some respects DEL induction and carcinogenesis may show similar characteristics. It may be one feature of nonmutagenic carcinogens that long term continuous exposure often at toxicologically significant doses is required to induce tumours⁶ which also seems to be true for DEL induction (Table 2). It also seems that nonmutagenic carcinogens may have a threshold in their dose response curve which may be missing with many mutagenic carcinogens²⁵. Similarly safrole (Fig. 2c), carbon tetrachloride, aniline, 3-aminotriazole, acetamide, ethylene thiourea, urethane (Fig. 2d) and thioacetamide all show thresholds below which no response can be detected. To the contrary the mutagenic carcinogens MMS (Fig. 2a), EMS, NQO, nitrogen mustard (Fig. 2b), dibromoethane, and aflatoxin increase the frequency of DEL recombination with increasing dose over a range of at least 10- to 1,000-fold (Table 1). Another similarity between carcinogenesis and DEL recombination was revealed when cells from Bloom's syndrome patients, a disease associated with a greatly increased incidence of cancer, were shown to contain an altered, heat-labile form of DNA ligase I²⁶. Similarly DEL recombination is induced about 100-fold in a yeast *cdc9* mutant temperature sensitive for DNA ligase¹⁷.

The result of this study may give an indication about molecular events involved in carcinogenesis. DEL events may give rise to recombination between repeated elements yielding genome

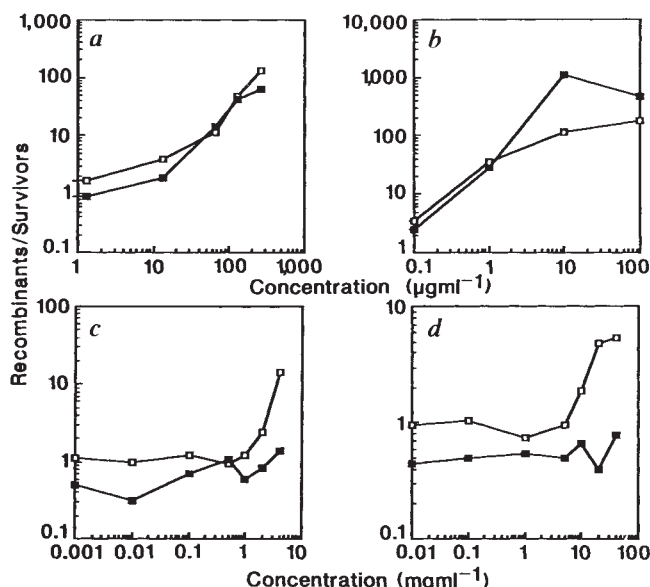


Fig. 2 Examples of dose-response curves for DEL ($\square$, events per 10^4 survivors) as well as ICR ($\blacksquare$, events per 10^5 survivors) recombination with two carcinogens which are detectable with the Ames assay, MMS (a) and nitrogen mustard (b) and two carcinogens which are not detectable with the Ames assay, safrole (c) and urethane (d). MMS and nitrogen mustard induce DEL as well as ICR in a dose dependent manner over a broad range of concentrations ($1,000\times$) whereas safrole and urethane induce only DEL recombination and show a threshold, below which no induction can be detected.

rearrangements and bringing proto-oncogenes⁸ into proximity with promoter or enhancer sequences. Naturally sister chromatid conversion may be important for the repair of damage induced in the S or G₂ phase and the nonmutagenic carcinogens may damage DNA specifically and only at this stage.

The DEL assay may exhibit more sensitivity to detect certain kinds of DNA damaging agents which are missed with the tests currently used. A large number of chemicals have to be tested to evaluate the DEL assay for its potential use in predictive carcinogenesis. This assay shows high reproducibility, is inexpensive, produces results within three days and requires only basic equipment and basic training in microbiology. If after further evaluation, results obtained with the DEL assay show high correlation with the carcinogenicity of the test agents, the DEL assay will be useful as a short term test for predictive carcinogenesis.

I am grateful to Professor U. Wintersberger at the Department for Tumour Biology, University of Vienna, under whose influence I developed the idea of constructing the DEL system. Dr R. D. Mehta of Prairie Biological Research Inc. is especially thanked for his help with the present work. Part of this study was carried out in the laboratory of Dr S. Prakash and was supported by Public Health Service grants from the NCI to S. Prakash. Drs R. D. Gietz, R. D. Mehta and Phil Hastings are thanked for many useful comments and for critical reading of the manuscript.

Received 21 September; accepted 23 November 1988.

1. Lave, L. B. & Omenn, G. S. *Nature* **324**, 29–34 (1988).
2. Ames, B. N., McCann, J. & Yamasaki, E. *Mutat. Res.* **31**, 347–364 (1975).
3. McCann, J., Choi, E., Yamasaki, E. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5135–5139 (1975).
4. McCann, J. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **73**, 950–954 (1976).
5. Flora, S. D., Zanaichi, P., Camoirano, A., Bemecelli, C. & Badolati, G. S. *Mutat. Res.* **133**, 161–198 (1984).
6. Sivak, A., Goyer, M. M. & Ricci, P. F. in *Nongenotoxic Mechanisms in Carcinogenesis*, 1–8 (Cold Spring Harbor Laboratory, New York, 1987).

7. De Serres, F. J. & Ashby, J. *Prog. Mutat. Res.* **1**, (1981).
8. Marx, J. L. *Science* **218**, 983–985 (1982).
9. Klein, G. *Nature* **294**, 313–318 (1981).
10. Pall, M. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2456–2468 (1981).
11. Cairns, J. *Nature* **289**, 353–357 (1981).
12. Wintersberger, U. *Naturwissenschaften* **69**, 107–113 (1982).
13. Schiestl, R. H., Igarashi, S. & Hastings, P. J. *Genetics* **119**, 237–247 (1988).
14. Ricci, P. F. & Melton, L. S. *Envir. Sci. Technol.* **19**, 437 (1985).
15. Hansen, M. F. & Cavance, W. K. *Cell* **53**, 172–173 (1988).
16. Ponder, B. *Nature* **335**, 400–402 (1988).
17. Schiestl, R. H. & Prakash, S. *Molec. cell. Biol.* **8**, 3619–3626 (1988).
18. *International Agency for Research on Cancer: Monographs on the Evaluation of Carcinogenic Risk to Man* Vols. 1–31 (IARC Lyon, 1973–1983).
19. Clapp, N. K. *Int. J. Cancer* **12**, 728 (1973).
20. Arcos, J. C. & Argus, M. F. *Chemical Induction of Cancer*, Vol. 11 A and B, (Academic, New York, 1974).
21. Zimmermann, F. K. *et al. Mutat. Res.* **133**, 199–244 (1984).
22. Ashby, J. & Tennant, R. W. *Mutat. Res.* **204**, 17–115 (1988).
23. *Handbook of Carcinogen Testing* (eds Milman, H. A. & Weisburger, E. K.) (Noyes, New Jersey, 1985).
24. Zimmermann, F. K. in *Cellular Systems for Toxicity Testing*, 186–196 (Annals of the New York Academy of Sciences, 407 New York, New York, 1983).
25. Kroes, R. in *Cellular Systems for Toxicity Testing*, 398–408 (Annals of the New York Academy of Sciences, 407 New York, New York, 1983).
26. Willis, E. A., Weksberg, R., Tomlinson, S. & Lindahl, T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8016–8020 (1987).
27. Fassullo, M. T. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6215–6219 (1987).
28. Sherman, F., Fink, G. R. & Hicks, J. B. *Methods in Yeast Genetics, A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1986).
29. *Nongenotoxic Mechanisms in Carcinogenesis* (eds, Butterworth, E. B. & Slaga, T. J.), (25 Banbury Report, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1987).
30. Shinozuka H., Lombardi, B., Sell, S. & Immario, R. M. *Cancer Res.* **38**, 1092–1098 (1987).
31. Sittig, M. *Handbook of Toxic and Hazardous Chemicals and Carcinogens* (Noyes, New Jersey, 1985).

Absence of significant RNA-dependent DNA polymerase activity in lymphocytes from patients with Kawasaki syndrome

Marian E. Melish*, Nyven J. Marchette†, Joan C. Kaplan‡, Susan Kihara‡, Donna Ching* & David D. Ho§

* Department of Pediatrics and † Department of Tropical Medicine and Medical Microbiology, University of Hawaii, John A. Burns School of Medicine, Honolulu, Hawaii 96826, USA

‡ Infectious Disease Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

§ Division of Infectious Diseases, Cedars-Sinai Medical Center and UCLA School of Medicine, Los Angeles, California, USA

Kawasaki syndrome, an acute febrile multisystem illness of young children, is a panvasculitis with prominent rheumatic features¹. Arthritis and pancarditis are frequent during the acute stage; coronary artery aneurysms occur in 20% of cases and the disease is now the leading cause of acquired heart disease in childhood. A microbial aetiology is suggested by the acute febrile self-limited character of the disease, the regular occurrence of epidemic outbreaks at intervals of 2–3 years, and the virtual restriction to young children, consistent with the early acquisition of immunity. Reports of elevated DNA polymerase activity (assumed to be RNA-dependent reverse transcriptase) in cultured lymphocytes from patients with acute Kawasaki syndrome suggest that a retrovirus might be the causative agent^{2,3}. We have measured supernatant DNA polymerase activity in lymphocyte cultures from 49 Hawaiian patients in acute and convalescent stages of Kawasaki syndrome and have been unable to demonstrate significant reverse transcriptase activity or other evidence of involvement of a retrovirus in the aetiology of the disease.

We established multiple cultures (two to four per patient) from peripheral blood lymphocytes of 36 patients with acute Kawasaki syndrome (KS) at intervals from days two to 40 or more of disease. An additional 13 patients were studied at one year or longer from onset of illness. Twenty-one KS patients

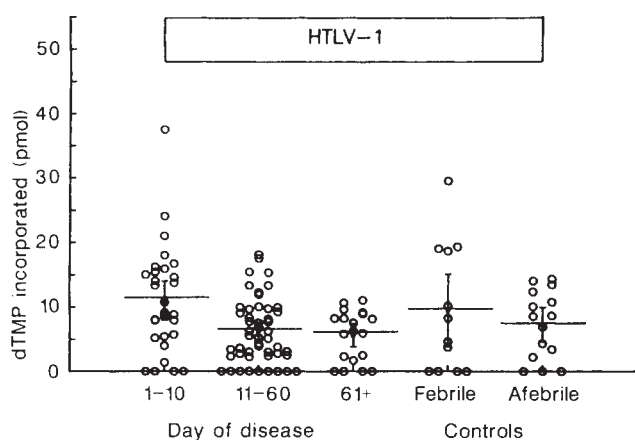


Fig. 1 Peak polymerase activity in the particulate fraction of supernatant fluids from cultures of lymphocytes collected from controls and KS patients at different stages of disease. Lymphocytes were isolated from peripheral blood of patients diagnosed as having KS by Dr Marian Melish using standard clinical criteria⁶. The mean age of the patients studied in the acute phase of disease was 29 months (range 4 to 94 months); that of control patients with febrile illness from infection with known or suspected viral or bacterial agents was 62 months (range 6 to 168 months); that of afebrile controls was 36 months (range 4 to 132 months).

Methods. Lymphocytes collected from heparinized blood by Ficoll-Hypaque sedimentation were cultured at a concentration of $1-2 \times 10^6$ per ml in Roswell Park Memorial Institute RPMI 1640 medium supplemented with 10% fetal bovine serum, glutamine, 2 mM final concentration, sodium bicarbonate sufficient to adjust final pH to 7.2, 100 units ml^{-1} each of penicillin and streptomycin, and phytohaemagglutinin ($10 \mu\text{g ml}^{-1}$). After 2-3 days incubation at 37°C in 5% CO_2 , the culture medium was decanted and replaced with complete medium supplemented with $0.5 \mu\text{g ml}^{-1}$ interleukin-2, 0.1% anti-human interferon- α , and $1.0 \mu\text{g ml}^{-1}$ Polybrene. After 5-8 days in culture, 3×10^6 Hut-78 or Molt-4 cells or activated fresh human lymphocytes were added to each 5.0 ml culture. Culture supernatant fluids were collected at 3-5 day intervals for 21-28 days and assayed for reverse transcriptase activity. On day 21, idoxuridine $100 \mu\text{g ml}^{-1}$ was added to selected cultures for 24-48 h, and then replaced with medium without idoxuridine. Supernatants were collected 4-5 days later. Reverse transcriptase assays were performed⁷. Briefly, culture supernates were centrifuged in a Sorval RC-2B angle-head rotor at 2,000 r.p.m. for 10 min at 4°C to remove intact cells and the supernatant fluid recentrifuged at 12,000 r.p.m. for 20 min to sediment cellular debris. The supernatant material was subsequently centrifuged at 25,000 r.p.m. for 90 min in an SW 28 rotor in a Beckman ultracentrifuge. In some cases, frozen (-70°C) pre-clarified culture supernatant fluids collected on days 3-9, 10-13, or 17-28 respectively were pooled before ultracentrifugation. The pellets were resuspended in 500 μl TNE buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 8.0), freeze-thawed twice, and a 10- μl aliquot was added to 10 μl reaction mixture containing 83 $\mu\text{g ml}^{-1}$ dATP, 50 mM Tris-HCl, pH 8.0, 5 mM dithiothreitol, 5 mM MgCl_2 , 50 mM KCl, 0.05% Triton X-100, 0.3 mM glutathione, 0.5 mM EGTA, 50 $\mu\text{g ml}^{-1}$ rADT, and 25 $\mu\text{g ml}^{-1}$ bovine serum albumin. After 5 min on ice, 7.5 μmol [^3H]dTTP were added and the mixture incubated at 37°C for 60 min with repeated shaking. Then 10 μl RNA (10 mg ml^{-1}) was added and the reaction stopped by the addition of ice-cold 0.01M sodium pyrophosphate and 10% trichloroacetic acid (TCA). The precipitate was collected on 24 mm Whatman GFA glass filters pre-soaked in 5% TCA. The filters were rinsed with chilled 0.01M sodium pyrophosphate in 1M HCl, dried, placed in scintillation cocktail (Acquasol-II, NEN), and counted in a Beckman beta scintillation spectrometer. Negative control (uninfected Hut-78 or Molt-4 cell) and positive control (HIV-1 and/or HTLV-I) culture supernates were included in all tests. Mean plus/minus 95% confidence intervals are indicated. The open bar across the top of the figure indicates 95% confidence intervals about the mean for cells infected with HTLV-I.

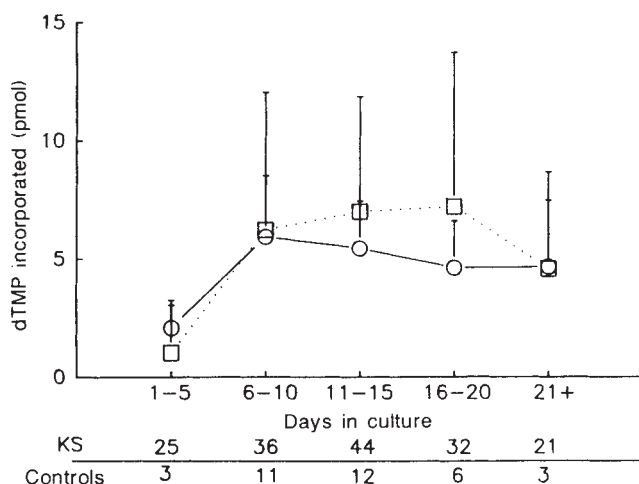


Fig. 2 Particulate polymerase activity in supernatant fluids of lymphocyte cultures. (○), KS patients in acute (febrile) phase; (□), febrile controls. The number of cultures tested at each time interval is shown on the horizontal line. Vertical brackets indicate 95% confidence intervals. Number of patients and control specimens tested at each time point is indicated below the graph.

were tested for serum antibodies to known human retroviruses—human immunodeficiency viruses types 1 and 2 (HIV-1, HIV-2) and human T-cell leukaemia viruses types I and II (HTLV-I, HTLV-II)—using the radioimmunoprecipitation assays previously described⁴. All patients except one were seronegative: one child with Japanese ancestry was seropositive for HTLV-I. DNA polymerase activity in supernatant fluid of lymphocyte culture was always low, with values significantly less than those of cell lines infected with HIV-1, HIV-2, and HTLV-I. The peak supernatant reverse transcriptase activity at every stage of the illness from the acute febrile period (from before day 10 after onset) to late convalescence (after day 30 from onset) did not differ significantly from that of febrile or afebrile child controls (Fig. 1). There was a tendency for polymerase activity in supernatant fluids from lymphocyte cultures from KS patients and febrile controls to be slightly higher after seven to 20 days in culture; but the mean was eightfold less than the mean for cells (MT-4, Hut-102) transformed by HTLV-I, a tightly cell-associated virus (Fig. 2). There was also little or no reverse transcriptase activity in cultures from afebrile convalescent patients (>10 days after onset) during any period of culture. In afebrile controls, reverse transcriptase activity was minimally elevated during the first four days of culture (Fig. 3).

In 13 cultures from 10 patients, DNA polymerase activity measured by dTMP incorporation was assayed using additional template:primers, poly (dA)·oligo (dT) and poly (rC)·oligo (dG), to determine whether the observed DNA polymerase activity was attributable to cellular DNA polymerases or to retroviral reverse transcriptase. In each case, the poly (dA)·oligo (dT) template:primer yielded values equal to those with poly (rA)·oligo (dT), suggesting that the low DNA polymerase activity was due to host cell enzymes. Furthermore, there was no activity detected with poly (rC)·oligo (dG), a template:primer preferred by RNA-dependent DNA polymerase⁵. Therefore, we were unable to demonstrate that the observed low-level DNA polymerase activity in supernatant fluids from lymphocyte cultures of KS patients was of retroviral origin.

Lymphocytes from five acute KS patients were also co-cultured with human peripheral blood lymphocytes, brain endothelial cells, and cell lines of T-lymphocytic (Hut-78, Molt-4, C8166, HPB-ALL), B-lymphocytic (Raji), and monocytic (U-937, BT4A3.5) lineage. In no case was RNA-dependent DNA

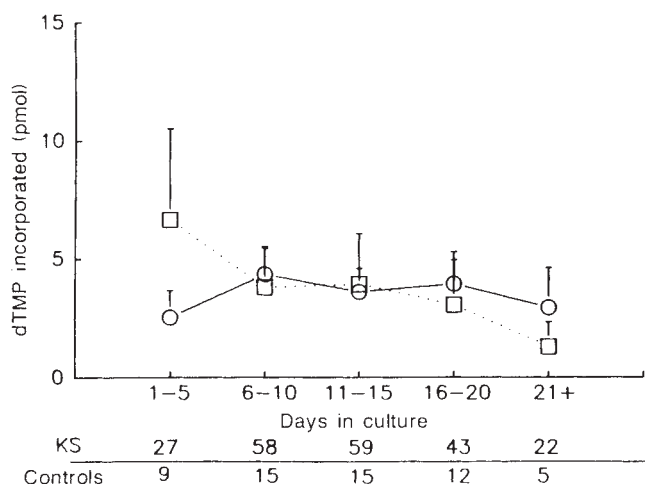


Fig. 3 Particulate polymerase activity in supernatant fluids of lymphocyte cultures. (○), KS patients in convalescent (afebrile) phase; (□), afebrile controls. The number of cultures tested at each time interval is shown on the horizontal line. Vertical brackets indicate 95% confidence intervals.

polymerase activity detected in the co-cultures. In one case however, syncytium formation was observed in a co-culture from HPB-ALL. Electron microscopic examination of this culture showed viral particles, subsequently identified as measles virus by specific immunofluorescence and radioimmunoprecipitation assays. The patient had been vaccinated with measles-mumps-rubella trivalent vaccine 8 days before onset of disease. It was concluded that the agent was measles vaccine virus.

Using standard methods for reverse transcriptase detection, we have been unable to demonstrate a retroviral aetiology for KS. The level and pattern of DNA polymerase activity detected in KS patients in our study were similar to those previously reported². In contrast to their study, cultures from our controls had values and patterns of activity similar to cultures from our KS patients. In addition, the low-level DNA polymerase activity detected in our cultures showed a template:primer preference consistent with cellular DNA polymerase and inconsistent with retroviral RNA-dependent DNA polymerase. Moreover, we have been unable to detect reverse transcriptase activity in repeated co-cultures of KS patients' lymphocytes with numerous human cell lines. Our experience with recovery of measles virus and its demonstration within cultured lymphocytes by electron microscopy emphasizes the danger in assuming that agents recovered from patients with KS may have a causative role.

Therefore, we believe that the aetiology of KS remains unknown. If a retrovirus is involved in the causation or the pathogenesis of this disease, techniques more sensitive than the assays presently available will be required to establish its presence.

We thank Drs Sandra Chang and Karen Yamaga for discussion and technical advice, Faith Caplan, Donna Ching, Terry Rota and Caroline Fong for technical support. Supported in part by the Hawaii Heart Association and by the NIH.

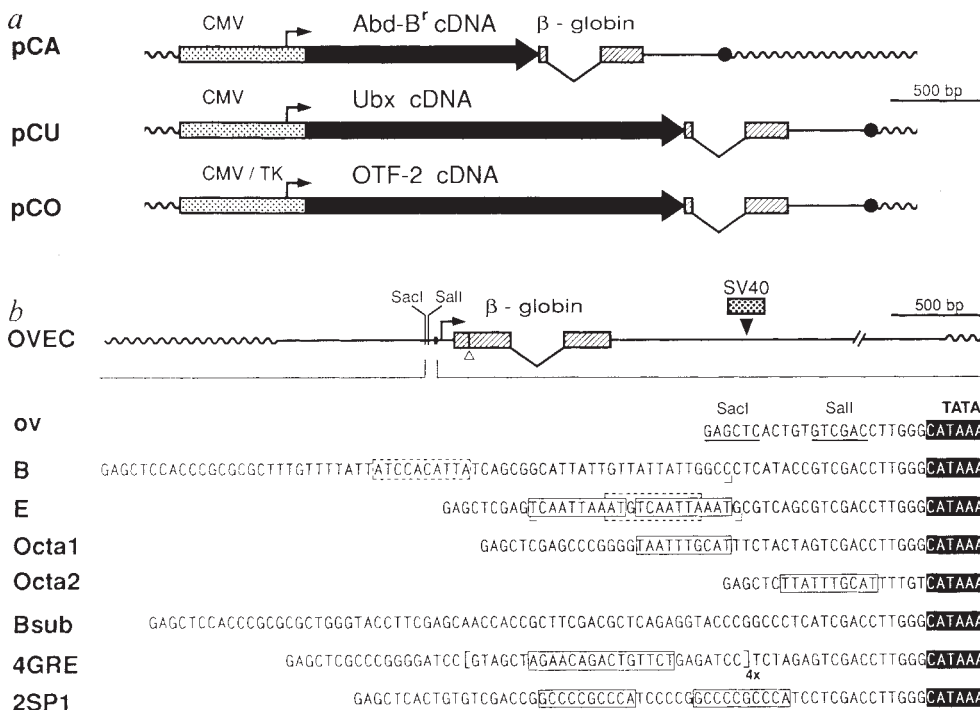
Received 12 September; accepted 11 November 1988.

1. Kawasaki, T. *Jap. J. Allergy* **16**, 178-222 (1967).
2. Burns, J. C. *et al. Nature* **323**, 814-816 (1986).
3. Shulman, S. T., Rowley, A. *Lancet* **ii**, 545-546 (1986).
4. Ho, D. D., Kaplan, J. C., Rackauskas, I. E. & Gurney, M. E. *Science* **239**, 1021-1023 (1988).
5. Rho, H. M., Poiesz, B., Ruscetti, F. W. & Gallo, R. C. *Virology* **112**, 355-360 (1981).
6. Melish, M. E., Hicks, R. M. & Larson, E. J. *Am. J. Dis. Child.* **130**, 599-607 (1976).
7. Hoffman, A. D., Banapour, B. & Levy, J. A. *Virology* **147**, 326-335 (1985).

Erratum: *Drosophila* homoeotic genes encode transcriptional activators similar to mammalian OTF-2

Markus Thali, Michael M. Müller, Mauro DeLorenzi, Patrick Matthias & Mariann Bienz

Nature **336**, 598-601 (1988)



IN this biological letter, an unrevised version of Fig. 1 was printed. The correct version is shown here.